Asbjørn Goul Andersen

Essays on inequality, health, and housing policy

August 2022

Dissertation for the Ph.D. degree

Department of Economics

University of Oslo
Acknowledgements

This thesis was written during my time as employed at The Ragnar Frisch Centre for Economic Research. I am grateful to the Norwegian Research Council for providing the PhD grant (project no. 258923) and to the Frisch Centre for welcoming me into an excellent research environment with friendly and inspiring colleagues who are genuinely interested and always ready to help.

I would like to thank several people for their contributions to this thesis. I am especially indebted to my supervisor, Andreas Kotsadam, who has guided and supported me, personally and academically, throughout the process of writing my thesis. You have been a great inspiration to me, and I thank you for always taking time out for me whenever I needed it. Writing the two last chapters of my thesis with you (among others) has been a very valuable learning experience. I am also grateful to my co-supervisor, Edwin Leuven, whose insights and suggestions have helped to improve my thesis.

I would further like to thank my co-authors on the first two chapters of my thesis, Simen Markussen and Knut Røed. It has been a great privilege to work with you both, and I enjoyed and learned a lot from our frequent meetings and discussions during my first years at the Frisch Centre. Likewise, I would like to thank Vincent Somville, who co-authored the last two chapters with Andreas and myself. It has been a pleasure to work with you, and I have appreciated your contagious enthusiasm, good sense of humor, and admire your motivation for getting things done today rather than tomorrow. Thanks also to Simon Franklin, Tigabu Getahun, and Espen Villanger, who coauthored the last chapter of this thesis with Andreas, Vincent and myself, for great cooperation.

Thanks are also due to Oddbjørn Raaum, Bernt Bratsberg, Ole Røgeberg, Viggo Nordvik, Kristin Aarland, and Jardar Sørvoll for valuable comments for the third chapter of the thesis. And thanks to Axel West Petersen for comments on the first chapter, and for not making me feel like the only Dane who is passionate about the Norwegian pension system.

Finally, a big thank you to Gry Nystrom, Jørg Gjestvang, Tao Zhang, and Tone Enger for always making things run smoothly.

This thesis is dedicated to my dearest Delali, Louie, and (forthcoming) Oscar.

Asbjørn Goul Andersen
Oslo, August 2022
## Contents

Introduction and [1]

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Pension reform and the equity-efficiency trade-off</td>
<td>27</td>
</tr>
<tr>
<td>2</td>
<td>Local labor demand and participation in social insurance programs</td>
<td>49</td>
</tr>
<tr>
<td>3</td>
<td>Settlement neighborhood crime and criminal behavior of young refugees in Norway</td>
<td>65</td>
</tr>
<tr>
<td>4</td>
<td>Material resources and well-being – Evidence from an Ethiopian housing lottery</td>
<td>133</td>
</tr>
<tr>
<td>5</td>
<td>Does wealth reduce support for redistribution? Evidence from an Ethiopian housing lottery</td>
<td>153</td>
</tr>
</tbody>
</table>
Introduction and summary
1. Introduction

This thesis consists of five essays revolving around the topics of inequality, health and housing policy. All essays rely on microeconometric methods to identify policy relevant causal effects. It has also been a general priority to ensure transparency regarding the research process in order to enhance credibility of the findings. This introduction is structured as follows. First, in this section, I briefly present the main content of the essays and how they relate to one another and to the overarching topic. Next, in section 1.1, I outline key elements of an ongoing debate about credibility in empirical microeconomics, before I discuss how such considerations have influenced the studies presented in this thesis (in section 1.2). Finally, the individual chapters are summarized in sections 2.1-2.5.

In the first chapter, we assess the distributional consequences of the elimination of a retirement earnings test in Norway in 2011. We identify causal effects of the reform by comparing the outcomes of adjacent cohorts, who faced very different work incentives after reaching the early retirement age of 62 years. While the transition to an unconditional pension increased income inequality directly, we find no evidence that unequal opportunities for postponing retirement in response to the reform amplified inequality further. In fact, the substantial labor supply response was strikingly homogenous and thus neutral in distributional terms. As documented by Hernæs et al. (2016), however, the reform increased disability insurance take-up among workers who lost an early retirement option. This underlines why policy evaluations should consider the potential spill-over effects to other parts of the social insurance system.

The second chapter applies such a ‘holistic’ perspective, when assessing how local employment opportunities affect inflows to different social insurance programs in Norway. To represent a source of exogenous variation in local labor demand, we use a “shift-share” instrument, which interacts the initial local industry-composition with subsequent national industry-specific employment fluctuations. In addition to the anticipated impact on the take-up of unemployment-related benefits, we find somewhat similar effects for temporary disability insurance. Our findings highlight the potential pitfalls of social insurance programs that make a sharp distinction between unemployment and disability.

The third chapter also deals with how unequal opportunities in the local environment may shape people’s life outcomes, albeit with a different focus. In particular, I exploit the quasi-exogenous assignment of refugee families to public housing units across Norway to investigate if children exposed to a criminal local environment are more likely to become involved in crime. I find no evidence that being settled in a higher-crime neighborhood increases criminal behavior. This may raise questions about the use of urban planning and housing policies to tackle social problems.
In the fourth chapter, we investigate how material living conditions affect mental wellbeing by studying winners and losers of a large-scale housing lottery in Ethiopia, which randomly allocates the right to purchase heavily subsidized state-built apartments. Winners experience a substantial wealth shock and report higher levels of overall life satisfaction and lower levels of financial distress than losers do. We find no effects on psychological distress, however, which appears to be less sensitive to economic conditions.

The fifth and final chapter uses the same Ethiopian housing lottery to investigate the causal effect of economic resources on political attitudes. We find no effects on general attitudes toward redistribution or inequality aversion, but new homeowners are less favorable of housing taxes. Interestingly, we also find evidence of endogenous beliefs: winners are more likely to attribute poverty to personality traits and less likely to emphasize the role of luck. This suggests that people may find ways to justify their improved economic status even if it is – as in our case – the result of a lucky draw. Such beliefs might (eventually) erode support for redistribution (e.g. Alesina and Angeletos, 2005; Almås et al., 2020), implying that rising inequality could potentially become self-reinforcing.

1.1 A note on credibility in empirical economics
The studies presented in this thesis all use quasi-experimental research designs to credibly identify policy relevant causal relationships. In that respect, they are in the spirit of contemporary empirical microeconomics. In this section, I will briefly reflect on this common denominator and relate it to ongoing discussions about credibility in empirical economics and in social science more broadly. First, I highlight some important improvements to the empirical practices within the discipline. I then briefly discuss remaining (and emerging) challenges and potential solutions and discuss how the empirical studies presented in this thesis have tried to overcome these problems.

As mentioned, the use of quasi-experimental methods is in many ways characteristic of empirical microeconomics; a discipline which, in the words of Angrist and Pischke (2010), has undergone almost a “credibility revolution” over the past decades. At the core of this development is the increased use of experimental and quasi-experimental studies and a more rigorous assessment of the assumptions required for the results to be comparable to those of a well-designed randomized trial. This “experimental ideal” increasingly serves as the benchmark against which the quality of a research design is judged.¹

¹ Randomized controlled trials ensure that assignment to receiving a given intervention (the “treatment” group) or a placebo (the “control group”) is uncorrelated with individual characteristics. As this eliminates
Actual randomization experiments have gained popularity in the form of lab- and field experiments, e.g. in behavioral- and development economics. While such designs generally require only weak assumptions, and provide causal estimates with high internal validity, they are often costly, and external validity may be poor. For instance, the findings from small (and often unrepresentative) samples in the “artificial” context of a lab, may not always generalize to the complex world outside the lab. Similarly, the results from field experiments based on e.g. extremely poor individuals in low-income countries may not apply to the country’s overall population or to other country-contexts, although this is arguably often not the main purpose of such studies either. Despite the high associated costs, there are also prominent examples of large-scale randomized experiments from high-income countries. These include the Moving to Opportunity (MTO) experiment, in which low-income families, living in deprived areas of five major US cities, were randomly assigned to receiving housing vouchers allowing them to move to better neighborhoods. As discussed in the third chapter of this thesis, evaluations of this experiment have challenged long-standing conclusions of theoretical and non-experimental studies about the importance of neighborhood effects.2

Besides the concerns regarding costs and external validity, randomized experiments are rarely useful for evaluating the impact of interventions or reforms that have taken place. Hence, for many policy-relevant research questions, experiments may neither be feasible nor informative. Instead, policy evaluations often rely on natural experiments or quasi-experimental research designs, in which individuals are exposed to different conditions in a way that is arguably uncorrelated with individual determinants of the outcome in question (under certain assumptions). The studies presented in this thesis fall into this category.

In addition to an increased scrutiny regarding the identifying assumptions, the availability of better data has greatly contributed to improving the quality of quasi-experimental research designs as well as the scope for applications. This is not least the case in Scandinavia where access to rich population-wide data from administrative registries has provided almost limitless opportunities for empirical inquiries (Røed and Raaum, 2003).

The advances made in empirical microeconomics are hardly disputable. However, the debate on credibility in social science is still flourishing. In an influential

---

2 More recent evaluations, have attributed the limited evidence of “neighborhood effects”, reported in earlier studies (e.g. Sanbonmatsu, 2011), to an insufficient account for heterogeneous treatment effects (Chetty et al., 2016). Others have questioned the premises for causal inference in a setting with multiple treatments, with reference to the two types of housing vouchers used in the experiment (Mogstad and Torsvik, 2022).
paper titled “Why most scientific research findings are false”, Ioannidis (2005) outlined a purely statistical framework for determining the likelihood that a given research claim (of a non-zero effect) is true. This so-called positive predictive value (PPV) depends – among other factors – on statistical power and “bias”.\(^3\) Even in the absence of any bias, lower power (i) increases the share of false negative findings, (ii) reduces the probability that an observed statistically significant effect represents a true effect, and (iii) exaggerates effects sizes when a true effect is discovered (Button et al., 2013).\(^4\)

“Bias” is defined as anything (other than chance) that may promote the finding, reporting, and eventual publication of false claims.\(^5\) Regardless of the cause, the presence of bias reduces the PPV and thus (further) weakens the credibility of published research claims. Simulations suggest that, at conventional levels of statistical significance, it may be hard to obtain a PPV above 0.5; hence the author’s bold claim that most published research claims are likely to be false.

A potential source of bias that have received much attention in this context is the possible favouring, among scientific journals as well as among researchers, for “strong” results, i.e. rejections of the null-hypothesis. In the case of journals, this often referred to as “publication bias”, and it might occur for various reasons. If the presence of an effect is considered “well-founded” (theoretically or empirically), editors and reviewers may be more likely to question the research design or data quality of studies that do not find an effect. On the other hand, for more speculative hypotheses, the existence of some “surprising” effect may simply make for a more interesting article than the absence of such an effect.

Researchers may prefer certain results, because they are in accordance with their prior beliefs or personal or commercial interests, or simply because they are expected to increase the probability of publication. Despite what is insinuated by commonly used terms like “data mining”, “p-hacking”, or “HARKing”, the bias toward certain types of results need not reflect deliberate manipulation or selective reporting. Rather, researchers may simply be less likely to question their model specification and data, when the results confirm what they already believe.

---

\(^3\) The framework also includes two other determinants; namely, the number of other studies of the same research question, and the ratio of actually true to false hypotheses among the hypotheses tested within a research field. Despite their theoretical importance, I refrain from discussing these components, because they are difficult to quantify. However, an implication of these two factors is that the credibility of research findings is likely to differ vastly across scientific fields and research questions. The PPV also depends on the chosen level of statistical significance, which is generally taken as given and is set to 0.05.

\(^4\) The first point follows from the definition of statistical power, while the second point follows from the expression for the PPV. The last point is a consequence of the fact that only results that are “extreme” (due to random variation in the data) will pass the set threshold for statistical significance.

\(^5\) In the following discussion, I use the term “bias” in this sense. Ioannidis briefly discusses the possibility of “reverse bias”, i.e. the (deliberate of accidental) “suppression” of true (non-zero) findings, e.g. due to conflicts of interest or simply inefficient use of the data, but argues that this is likely to be of lesser practical importance.
The described types of bias have two important implications. First, the results of published studies may not be representative of the results obtained in all (equally qualified) inquiries into the same research question. Second, even if established findings may be challenged in new studies, science may not always be “self-correcting”.

Ioannidis’ (2005) article gave rise to much introspection within several scientific fields, where it may have offered an explanation for what has been called a “replication crisis”, referring to the fact that many – and even ‘well-established’ – scientific findings have often proven difficult to replicate. It appears to also have reignited the old debate on credibility in economics (e.g. Leamer, 1983; De Long and Land, 1992). Indeed, some of the challenges outlined by Ioannidis (2005) seem highly relevant for empirical economics. First, the topic of statistical power has traditionally received little attention among economists (Ziliak and McCloskey, 2004). Second, while replication studies and meta-analyses play a prominent role in many other disciplines, these may be somewhat less regarded in the economics literature, which appears to have a stronger preference for novelty. Finally, the common use of observational data with many variables and few standards regarding the exact empirical specifications, may allow for a large degree of flexibility in the research process.

According to Ioannidis et al. (2017), evidence seems to support such concerns. Based on a survey of meta-studies, they conclude that the empirical economics literature suffers from severe power and bias issues, and that most of the reported effects are likely to be highly exaggerated. However, meta-analyses are not as common in economics as in other scientific fields, and the included studies may not be representative of the entire discipline.

There is also evidence of specific types of bias in empirical economics. Franco et al. (2014) study a population of published and unpublished studies, whose research designs had been all been subject to a similar peer-review process before the researchers were given access to data. They find that null-results are much less likely be published than are “strong” results. Indeed, much of this “publication bias” occurs even before the actual publication process: studies finding null-effect are much less likely to be even written up (the so-called “file-drawer problem”).

Brodeur et al. (2016) examine the distribution of z-statistics (the “z-curve”), of results reported in highly ranked economics journals, and find a double-humped shape, which cannot solely be explained by publication bias. The authors interpret the apparent “deficiency” of marginally rejected tests as evidence that researchers, who initially find such results, may adjust their specifications in order to pass the threshold for statistical significance.

---

6 It is even considered as founding for the entirely new discipline “metascience”, where scientific methods are used to study science itself. Indeed, some of the studies discussed below may be seen as belonging to this discipline.
The above discussion suggests that despite the great improvements made in empirical economics, important challenges remain. In the next section we discuss some of the suggested remedies and how such considerations have influenced the studies presented in this thesis.

1.2 Credibility considerations in this thesis

In the empirical studies making up this thesis, I (and my coauthors) have taken different steps to improve transparency regarding the research process in order to strengthen credibility of the findings. I outline these below in the context of a more general discussion of some of the proposed strategies for improving the quality of empirical research.

As mentioned, the lack of standards or strong priors regarding the “true” relationship implies that the chosen empirical specification may often be selected among many possible specifications. The “classic” solution to this problem is to perform robustness checks or a sensitivity analysis, as suggested by Leamer (1983). We follow such an approach in chapter 2, where we present a version of our results for five alternative specifications. The fact that these rather different models yield almost identical results may provide some reassurance that our main results are not cherry-picked or an artifact of the empirical specification. To rule out the latter possibility, we also perform a placebo-analysis, where we replace the outcome variable with a variable that should not be affected by the explanatory variable. To be more specific, instead of testing whether the variation in local labor market conditions affects future labor market states, we test whether it affects local labor market states in the past. Again, it is reassuring that we find no effects for this alternative outcome.

Robustness checks chosen by researchers may, however, not always rule out cherry-picking. For this reason, referees may often request specific robustness analyses. An alternative solution that has recently gained popularity, is to produce estimates for a wide range of combinations of empirical decisions – usually by means of some automated process – and then rank them and present all the results together. In addition to demonstrating the sensibility of the results to the empirical specification, such so-called “specification curve analyses” (Simonsohn et al., 2020) might provide insights about which control variables are important and which may be irrelevant.

The increasing availability large and rich dataset, including so-called “big data” from electronic sources, provides many new interesting research opportunities. However, such large datasets and the lack of good priors may also increase the scope of possible specifications or for selective subgroup analyses. Wager and Athey (2018) argue that, in such cases, outsourcing the task of picking the best specification to a
machine learning algorithm might enhance credibility of the findings.\textsuperscript{7} Such methods are essentially computerized data mining, but by using separate samples for fitting and testing the model, the authors argue that it is possible to obtain “honest” estimates.

However, machine learning methods have also gained popularity outside “big data” settings. In the two last chapters of this thesis (i.e. the housing lottery studies), we present a version of our results where we include a set of “optimal” control variables selected using the “post-double lasso” procedure suggested by Belloni et al. (2014). Moreover, in chapter 4, we explore the potential heterogeneity in the treatment effects using the “generic machine learning approach” of Chernozhukov et al. (2018). Even though none of these methods provided insights different from those conveyed by our main results (which is also somewhat reassuring), they do represent new innovative strategies for reducing selective reporting.

While machine learning methods might improve credibility in some situations, the results from such methods are sometimes hard to interpret and may not be useful for policy design. Also, many decisions are still left to the researcher, including the choice of method and input variables, or selection of the sample). Hence, there is still room for developing methodological standards or conventions in this area.

Another way to improve credibility, may be to ensure more transparency regarding the research process (Christensen and Miguel, 2018). This may be achieved in different ways and many of the suggested strategies have already been widely adopted. For instance, many journals now encourage (or require) researchers to share their data and code-files when an article is accepted for publication. Apart from disclosing the decisions made in the handling of the data, this allows other researchers to reproduce or replicate the presented findings and assess their robustness. If possible, we plan to share our data and code-files in relation to the publication of the paper presented in chapter 5.

Data-sharing may, however, not always be possible. For instance, the Norwegian registry data used in the first three chapters may not be shared for privacy reasons. On the other hand, for studies based entirely on data from public registries (i.e. not matched with data from other sources), it may in principle often be possible for other researchers at Norwegian research institutions to get access to the exact same data in relation to other research projects. In practice, however, access to registry data is both expensive and strictly regulated, but new initiatives such as “microdata.no”, allows a broader population of students and researchers (though still only at Norwegian research institution) to perform analysis on administrative data without actually

\textsuperscript{7} The authors also argue that the lack of good priors may make it difficult to pre-specify the most relevant hypotheses. Hence, pre-analyses plans may be too restrictive and leave much heterogeneity uncovered.
accessing it. Hence, even if the data cannot be shared, the sharing of code-files may still be valuable.  

There may also be other ways of overcoming the restrictions pertinent to registry data. Indeed, in the study presented in chapter 2, we develop such a method. Because aggregating data at the commuting-zone-by-year level yielded results almost identical to those based on individual level data (as shown in the appendix), we chose to report these results in the main text. In addition to being computationally more efficient, this would enable us to share our data publicly, allowing others to replicate and test the robustness of our results.  

Hence, for analyses based on e.g. geographical units, such as that of chapter 3, aggregating individual-level data may provide a solution to the problem that registry data cannot be shared due to privacy concerns.  

Transparency may also be improved through pre-registration. Pre-registrations have also gained popularity, and some journals even commit to publishing studies based solely on the pre-analysis plan. This in turn has heightened demands for proper power calculations in advance of the study. As pointed out by Button et al. (2013), preregistration of hypotheses and empirical specifications prevents researchers from presenting exploratory findings as confirmatory findings. It also ensures, that \( p \)-values have a meaningful interpretation.  

The studies presented in the two final chapters of this thesis were based on a pre-analysis plan which specified the main hypotheses, the empirical specification, the coding of variables, and a series of robustness checks and secondary outcomes. It also stated that we would adjust the critical \( p \)-values for multiple hypothesis testing across two-three different papers using the false discovery rate method proposed by Benjamini and Hochberg (1995). Critics have argued that pre-registration may discourage exploratory analyses and careful robustness checks. As mentioned, we described many such analyses in the pre-analysis plan; however, there was nothing stopping us from doing additional analyses that either we or the referees thought might provide new insights. One of the interesting findings in chapter 5 was on an outcomes that, even though it was included in the pre-analysis plan, was not specified as a main outcomes. To be more specific, this relates to the observation that lottery winners were less likely than losers to emphasize the role of luck when asked about the causes of poverty. In order to give this finding more credibility, we use data from a survey of winners and losers from an earlier round of the lottery, which included questions similar to ours.

---

8 Researchers who have invested much time and energy in the preparation of datasets that provide them with a “comparative advantage” relative to other researchers, may of course be reluctant to share their code-files.

9 The intention was to publish the data and code-files a long with a short “Data-in-brief”-article, however, it turned out to be slightly more of a hassle than initially expected, and we eventually gave up. A recommendation based on this experience may be that data-sharing should be easy.
The results from this study were consistent with our findings both for this and other outcomes.

Based on simulations of the framework of Ioannidis (2005), Coffman and Niederle (2015) find that pre-analysis plans may have a limited upside in terms of improved credibility (at least for certain parameter values), especially if they are not extremely detailed.\textsuperscript{10} Their simulations suggest that promoting replication studies may be a more appropriate strategy for improving credibility.\textsuperscript{11} Replication studies are, however, found to be particularly valuable when bias is low, which is why the authors recommend (somewhat ironically) that this particular type of studies should be pre-registered.

Coffman and Niederle (2015) also argue that pre-analysis plans may in fact be more important for observational studies than for the experimental studies, where they are currently used the most. They base this claim on the fact that the two-hump shape in the z-curve, documented Brodeur et al. (2016), was more pronounced for the observational studies than for the experimental studies. As discussed, a possible explanation for this is the multiplicity of variables often available for such studies. It is not obvious, however, how to credibly pre-register an analysis based on existing data that may have been accessed previously by one or more members of a research team, as is often the case for registry data. One solution to this problem may be to prespecify an analysis to be carried out on data that is currently not available but will become available in the future. This might apply to entirely new datasets that are under preparation, e.g. by Statistics Norway, or to updates of existing datasets. Indeed, such approaches have already been followed. For instance, Clemens and Strain (2021) developed an estimation framework for studying the effects of minimum wage changes based on a short-term analysis (from 2011-2015), and committed to using the same framework for a medium-term analysis (until 2019) as more data would become available. However, given the uncertainty regarding data-quality and anything else that might happen (in the world or in the life and careers of researchers), this solution may have its limitations.

For the study presented in chapter 3, I have chosen a simpler approach: I developed a pre-analysis plan based on data that was currently available to me for the analysis of that same data. The plan, which was registered at the EGAP-registry (Reg. ID: \texttt{20220513AA}), specifies the main hypotheses, empirical specifications, sample selection, operationalization of concepts and variables, robustness checks, and some

\textsuperscript{10} The authors also argue that while hypothesis-registries may allow for a better assessment of the credibility of published findings (as well as that of researchers), they might be of little value in practice. Even if this conclusion might seem a bit too pessimistic, I shall not discuss hypothesis-registries further here.

\textsuperscript{11} Interestingly, the authors' proposal for a dedicated journal for replication studies appears to have been met, as exemplified by the newly launched "Journal of Comments and Replications in Economics".
suggestions for additional exploratory analyses. By coding the variables in advance, I avoided unpleasant surprises regarding data quality, and I was even able to perform a balance test in order to validate my empirical design. In fact, the only thing I strictly refrained from doing was to merge the outcome data with the dataset containing the main explanatory variables and the covariates.

This suggested approach is somewhat unconventional in that it is entirely trust-based. In this light it seems reasonable the question whether it offers any improvements relative to studies that are not pre-registered. If the main sort of bias in conventional research is due to deliberate manipulation, this type of pre-analysis plan may be of little practical value. However, if one believes that most researchers are honest people, who nevertheless may be subject some sort of “confirmation bias” – making them more likely to question their empirical design, or to leave their findings in the file-drawer, when results at odds with their prior expectations – there might be a case for this type of pre-analysis plans. Indeed, it may even have some advantages relative to more conventional pre-registration procedure. Apart from enabling a validation of the empirical design beforehand, familiarity with the data and its limitations makes it much easier to develop a plan with a high degree of detail, without having to specify too many solutions for hypothetical situations where the data might prove inadequate. In this sense, it may be less of a hurdle, and less prone to error than conventional pre-analysis plans.

My pre-registration of the register-based study was done somewhat as an experiment, inspired by discussions with colleagues at the Frisch Centre about the topic, and I can only speak from my own experience. While it has been argued that pre-analysis plans are too constraining, I actually felt that developing and following the plan was somewhat liberating. It also made me deeply aware of the seemingly endless number of decisions where one option may be as reasonable as another, and that my pre-registered main specification is truly drawn from a distribution of many possible specifications. Worse yet, it might not be the “best” or most correct specification. However, I have not felt too constrained in the process of writing the paper. After all, it is possible to deviate from the plan, although such deviations should be reasonably justified. Also, nothing has prevented me from doing all the exploratory analyses I could think of, but I would of course have to present such analyses as what they are; namely exploratory.

In the above discussion of transparency and credibility, I have tried explain how such considerations have been central to the studies presented in this thesis. In the next section, I provide a brief summary of each of the individual chapters.
2. Chapter summaries

2.1 Chapter 1: Pension reform and the equity-efficiency trade-off

Co-authored with Simen Markussen and Knut Roed

In order to address the rising fiscal costs related to population aging, many countries are seeking ways to encourage later retirement. One such strategy may be to reduce the implicit tax rates on income from work beyond the statutory retirement age by relaxing or removing the earnings test on retirement benefits. In a pension scheme with no deferral option, a retirement earnings test (RET) often entails high implicit tax rates on work beyond the retirement age, discouraging continued employment even when its social value exceeds the private value of the forgone leisure. However, by effectively transferring pension wealth from (often high-income) late-retirees to (often low-income) early-retirees, RETs also promote a more even distribution of old-age income. Hence, although RET-reforms may improve economic efficiency, they are also likely to increase inequality directly. Moreover, if workers have unequal abilities to respond to strengthened work incentives (Etgeton, 2018), this might amplify (or in principle offset) the direct distributional effect. While, the potential efficiency gains from RET-removals are well-documented (Baker and Benjamin, 1999; Brinch et al., 2015; Brinch et al., 2017; Hernæs et al., 2016), existing evidence on the distributional effects of such policies is scarce and somewhat ambiguous (Hernæs and Jia, 2013; Bönke et al. 2018).

In this article, we investigate the tradeoff between efficiency and equity in the context of a Norwegian pension reform from 2011, which effectively abolished an early retirement earnings test for about half of the private sector workforce still active at age 60. The main empirical challenge in determining the impact of the reform is to find a good control group whose labor market behavior may serve as a credible counterfactual for the behavior of the workers targeted by the RET-removal. This task is not straightforward, however, given that the reform included other elements, and that most private sector workers were affected in one way or another. Public sector workers, on the other hand, may be different from private sector workers in many ways and may be seen as operating in a different labor market.

Instead, we exploit that the reform was implemented in a quasi-experimental manner, which entailed that adjacent cohorts faced vastly different incentives for work beyond the early retirement age of 62 years. To be specific, we identify the causal impact of the RET-removal by comparing workers of the last pre-reform cohorts to those in the first post-reform cohorts. We argue that the income trajectories of these
cohorts would have been largely identical in the absence of a reform, after controlling for any observable differences. This argument relies on three main identifying assumptions: (i) no selection into or out of the group targeted by the RET-removal, (ii) no cohort or calendar time effects, and (iii) no direct or indirect effect of the reform on the pre-reform cohorts (i.e. the control group).

Using detailed administrative data covering the entire Norwegian population, with employer information and individual earnings histories from 1967 onwards, we are able to single out the group of private sector workers affected by the RET-removal. To minimize the risk of selection into or out of the group targeted by the RET-removal, we base group assignment on the employment status at age 60 (the age of the youngest cohort when the reform was announced). Cohort or calendar time effects appear to be of minor importance: the pre- and post-reform cohorts have very similar earnings and employment histories, and there are no trends in labor supply after age 62 within the pre- and post-reform cohorts, respectively.\(^1\) By contrast, the difference in labor supply \textit{between} the pre- and post-reform cohorts is substantial. Finally, spillover effects of the reform on the pre-reform cohorts – e.g. due to competition for a limited number of jobs or related to joint retirement in couples – cannot be ruled out, but we argue that they are likely to be negligible.

We carry out an empirical analysis in three parts. First, we first reproduce and extend the results of Hernæs et al. (2016) by showing that the initial labor supply effects remained strong until the age of 67 years. Moreover, the response was strikingly homogenous across the distribution of previous earnings as well as along other dimensions; in particular, personal sick-leave history and occupation based life-expectancy and social status.

Second, we characterize the definite winners and losers of the transition from an earnings tested pension to a lower unconditional pension. Under the assumption that nobody decides to leave (remain in) employment as a result of higher (lower) take-home wages, we can describe these groups directly from the data. In comparison to losers, the winners have higher prime-age earnings, higher education, more prestigious occupations, and much lower sickness absence in the past.

Finally, we examine the impact of the reform on the distribution of accumulated pension and labor income after the age of 62 years in a way that allows us to disentangle the direct effect of the new entitlement rules from the consequences

\(^1\) This is consistent with the relatively stable macroeconomic environment during the period considered. A common way to control for cohort and calendar time effects is by using a difference-in-differences setup. However, given that all private sector workers were affected by the reform to some extent, public sector workers (qualifying for the early retirement program) constitute the only candidate for a control group. Even if it were possible to demonstrate parallel pre-reform trends between the two groups, public and private sector workers are arguably different and operate in different labor markets. Hence, such an approach would likely introduce more problems than it would solve.
of the reform-generated changes in labor supply. We show that the new entitlement rules entailed a considerable increase in old-age income inequality, whereas the homogenous labor supply response was almost neutral in distributional terms.

Our finding should be interpreted in light of the relatively resourceful group of workers targeted by the RET-removal. We essentially limit attention to workers with relatively stable employment and earnings histories who are still employed at age 60. Although this group may be quite representative for the type of workers that governments might successfully convince to extend their labor market careers, they are not representative of the entire population of senior workers. Also, our analysis focuses purely on the monetary aspects of inequality and implicitly assigns leisure a value of zero (for everyone). However, the retirement behavior of a transition cohort, who could freely choose between the old and the new scheme, indicates that workers at the lower end of the income distribution may incur a higher utility-cost from continued work.

Even if our results may understate the adverse effects on inequality, e.g. in “quality of life”, the observed distributional impact of the RET-removal is substantial. Nevertheless, in light of the considerable efficiency losses associated with RETs, there might be less costly ways of achieving a more equal distribution of old-age income.

2.2 Chapter 2: Local labor demand and participation in social insurance programs

Co-authored with Simen Markussen and Knut Roed

Social insurance programs generally draw a sharp distinction between unemployment and disability; however, in practice, the causes underlying social insurance claims may often be ambiguous, and there is evidence of some program substitution (e.g. Autor and Duggan, 2003; Maestas et al., 2013). Studies have even documented a causal effect of employment prospects on disability claims from employment shocks, e.g by exploiting employment shocks related to fluctuations in natural resource industries (Black et al., 2002; Charles et al., 2018) or to plant downsizings or closures (Rege et al., 2009). Moreover, the size of such effects appear to depend on the local labor market conditions (Bratsberg et al., 2013).

In this article, we explore the “gray area” between unemployment and disability, by investigating how local employment prospects affect entrance to different social insurance programs. The challenge in this context is to find a good proxy for local labor market conditions that is plausibly exogenous. We obviously cannot use the local employment rate, which is determined by the intersection between local labor demand and supply. To overcome this “simultaneity problem”, we follow “shift-share”
instrumental variables strategy, in which we interact the initial local industry-composition with the subsequent nationwide industry-specific employment fluctuations, in order to obtain a source of exogenous variation in local labor demand.

Our analysis offers two novel contributions. First, unlike studies of employment shocks in specific firms or industries, our analysis includes the entire working age population. Hence, external validity is likely to be high. Second, by assessing the effect of local labor market conditions on entrance to different social insurance programs, we can directly compare the impact on disability-related insurance claims to the effect on unemployment-related insurance claims.

We find that better (worse) local employment conditions reduce (increase) take-up rates for not only unemployment-related insurance, but also for disability-related insurance. For newly employed workers, the effect on disability-related insurance is nearly one-third of the effect on unemployment insurance. The effects are larger, and even more similar for newly unemployed workers. Our findings are robust to alternative specifications of the shift-share instrument and the empirical model. Moreover, a placebo analysis, in which we substitute the future labor market outcomes with past labor market outcomes, suggests that the results are not an artifact of the empirical specification.

There may be several explanations why local employment conditions might affect disability insurance claims, including caseworker leniency in tough times, and the potential health effects of (un)employment. Regardless of the cause, the different social insurance programs involve different follow-up strategies, and therefore program assignment is likely to have real consequences for long-run labor market outcomes (Schreiner, 2019). Hence, our findings suggests that the sharp distinction between unemployment and disability embedded in many social insurance programs may not always be appropriate.

2.3 Chapter 3: Settlement neighborhood crime and criminal behavior of young refugees in Norway

Single authored

“Neighborhood effects” have traditionally been considered important in the context of youth criminal behavior (Case and Katz, 1991; Sampson, 2002). However, causal evidence is in fact more mixed than often assumed (e.g. Kling et al., 2005; Damm and Dustmann, 2014). Determining the causal impact of growing up in a high-crime neighborhood, because certain kinds of people may sort into certain types of neighborhoods. A simple comparison of people growing up in areas with different crime
rates will therefore likely be subject to selection bias. In this paper, I provide new causal evidence on this topic focusing on a high-risk population; namely refugees who are offered resettlement in Norway.

My empirical design exploits the fact that refugee children are quasi-exogenously exposed to local environments with vastly different crime rates upon arrival to Norway. The absence of prior contact between the refugees and officials in charge of the settlement process greatly reduces the risk of selection based on unobserved behavioral characteristics. This allows me to identify the causal impact of the settlement neighborhood.

I find no evidence that being settled in a high-crime neighborhood increases youth criminal behavior. This conclusion holds for various types of crime and regardless of the geographical unit used in the analysis. Moreover, based on the results from a pre-registered main specification, I can reject relatively small effect sizes.

While other studies have used refugee dispersal policies in to analyze neighborhood effects on youth crime (e.g. Damm and Dustmann, 2014; Grönqvist et al., 2015), the present study offers several new contributions. First, the refugees in this study are assigned to specific housing units (and neighborhoods), before even arriving in Norway. This reduces the risk of selection problems relative to studies of e.g. asylum seekers (Nekby and Pettersson-Lidbom, 2017). To further strengthen identification, I impose strict conditions on the prior arrival of other family members and carefully exclude refugee families for whom the initial assignment is unlikely to be exogenous.

Next, the availability of high-quality geographical data, allows me to characterize neighborhoods at a very local level and to assess the sensitivity of the results to the chosen level of aggregation.

Also, the fact that all main hypotheses, model specifications, and operationalizations of variables and concepts were pre-registered, before performing any of the analyses, may enhance the credibility of the empirical findings.

Finally, I believe that my results may have a clearer interpretation than those of studies attempting to “isolate” the effect of local crime by controlling for various characteristics of the local environment. My empirical design aims to identify the effect of exposure to a neighborhood with a higher crime rate upon arrival to Norway, regardless of the underlying causes and correlates of this crime rate. This may be more informative and policy relevant than e.g. statements about the impact of local crime for a given level of poverty or immigrant share.

Interestingly, my conclusions differ from those of an influential and rather similar study from Denmark by Damm and Dustmann (2014). While this might be attributed to differences in the level of aggregation, population, country context, and time period considered, the most important explanation may in fact be that the causal parameters of the two studies have different interpretations. Indeed, their results from
a model more similar to the one used in this study are consistent with the findings presented above. As argued above, I believe these estimates have a clearer and more policy-relevant interpretation.

My findings are, however, consistent with and add to a causal literature, suggesting that neighborhood effects in crime may not be as important as traditionally assumed (e.g. Sciandra, 2013). They are also consistent with evidence from neighborhood-correlations, indicating limited scope for neighborhood influences in criminal behavior and other outcomes, whereas family background is very important (Eriksson et al., 2016; Raaum et al., 2006). In this light, one might ask if the widely adopted policies, focusing on changing the residential composition of disadvantaged neighborhoods, to some extent may be missing the target.

2.4 Chapter 4: Material resources and well-being – Evidence from an Ethiopian housing lottery

Co-authored with Andreas Kotsadam and Vincent Sommeville

While the positive correlation between economic resources and mental well-being is an almost universal finding (e.g. Deaton et al., 2008; Diener et al., 2010; Killingsworth, 2021), the nature of this relationship remains poorly understood. For instance, poverty might be both a cause and a consequence of poor mental health; a topic of increasing concern in low- and middle-income countries today (Haushofer and Fehr, 2014; Ridley et al. 2020).

In this article, we investigate the causal effect of better material living conditions on life satisfaction and the prevalence of common mental disorders. We survey roughly 1500 winners and 1500 losers of a large-scale housing lottery in 2016, which allocated the right to purchase a newly state-built apartment in Addis Ababa, Ethiopia, at a highly subsidized price. Because winning the lottery is random, we can identify the causal effect of winning by comparing the responses of winners and losers.

Our results confirm that winners experience a substantial wealth shock. They also report higher levels of overall life satisfaction and lower levels of financial distress. We find no effects on psychological distress, however, and our point estimates allow us to reject relatively small effects (0.1 standard deviations).

Our results are in contrast to those of many other studies from low-income countries, finding significant mental health effects of much less dramatic economic interventions, e.g. cash transfers to very poor individuals (see e.g. McGuire et al., 2020). Interestingly, our findings are in line with those of a large lottery study from Sweden, in which the authors attribute the absence of mental health effects to the
comprehensive Swedish social security system (Lindqvist et al., 2020). In this light, it seems remarkable to find similar results in an Ethiopian context. A possible explanation is that even though the lottery participants in our study are poor, they are not among the poorest of Ethiopians. In order to be eligible for the lottery, registrants must open a bank account and deposit savings every month, which requires a relatively stable and secure financial situation.

Interestingly, we find that winning has the most beneficial effects on well-being for individuals belonging to second quartile of the earnings distribution. This may be explained by the character of the “intervention”. While winners undoubtedly become much wealthier, they also have higher expenditures (at least initially) in the form of a down-payment on the apartment, and the monthly mortgage payments. Despite exhibiting lower levels of financial distress on average, winners have more debt and report more difficulty in raising additional funds for unforeseen expenditures.

Another difference between our context and those of most cash transfer studies, is that lottery winners simultaneously become wealthier and obtain access to better housing. While, we cannot fully disentangle the effects of these changes, only 30 percent of winners had moved into their new apartment at the time of the interview, and we find similar effects on all well-being outcomes for movers and non-movers. Although the decision to move is not random, and hence comparing the two groups involves a risk of selection bias, the similarity suggests that wealth rather than housing conditions is the main driver of our results. This is confirmed by a mediation analysis, which shows that the majority of the observed effect on satisfaction is mediated by wealth, whereas there is no evidence of the effect being mediated by moving.

Our findings are consistent with the notion that life satisfaction and mental health may have different determinants (e.g. Kahneman and Deaton, 2010). Many low-income countries (increasingly) use social housing programs as a means of improving the living conditions of their citizens; however, our results suggest that economic security might be of greater importance for overall life satisfaction than housing conditions are. Moreover, our results suggest that improving mental health likely requires other and more specialized policies than merely improving people’s material living conditions.
2.5 Chapter 5: Does wealth reduce support for redistribution?
Evidence from an Ethiopian housing lottery

Co-authored with Simon Franklin, Tigabu Getahun, Andreas Kotsadam, Vincent Sonville, and Espen Villanger

Correlational evidence suggest that wealthier people are less supportive of redistribution (Alesina and Giuliano, 2011); however, the extent to which this reflects a causal effect of wealth rather than just selection has rarely been examined. Out of self-interest, wealthier people might oppose redistribution, because they are more likely to have to pay for it. However, support for redistribution may also depend on values and fairness considerations (Cappelen et al., 2007). Moreover, people are more likely to consider income differences as fair if they result from effort rather than luck (Almås et al., 2020).

Determining the causal effect of material resources on attitudes is challenging, because more and less affluent people are likely to be different in many other regards. In this article, we provide new causal evidence on this topic from a pre-registered survey of winners and losers of an Ethiopian housing lottery. The lottery randomly allocates the opportunity to purchase an apartment at a highly favorable price, and winners thus experience a substantial wealth gain. Because winners are randomly drawn from the pool of lottery participants, we can identify the causal effect of winning the lottery by comparing the attitudes of winners and losers.

We find no effects on general attitudes toward redistribution or inequality aversion, indicating that such attitudes may be rooted in deep and relatively stable values. New homeowners are, however, less favorable of housing taxes, indicating that support for particular redistributive policies may be driven by self-interest. Interestingly, winning the lottery also appears to change people’s beliefs about the sources of income inequality: Winners are more likely to attribute poverty to personality traits (“poor character”) and less likely to emphasize the role of luck. This suggests that people may (unconsciously) find ways to justify their economic status even if it is — as in our case — a result of a lucky draw.

The question about the role of luck in determining economic outcomes was not one of our pre-registered main outcomes. In order to enhance credibility of this finding, we therefore replicate the results using survey data collected by Franklin (2019) from a previous round of the lottery, which included similar question. These results are consistent with the findings presented above, both with respect to the belief about the role of luck and as regards the absence of an effect of winning on general preferences for distribution. The fact that the observed changes in beliefs do not immediately lead to changes in support for redistribution may seem surprising. It might reflect that distributional preferences are more stable than beliefs (Fisman et al., 2020) and might take longer to adjust.
References:


Schreiner, Ragnhild. 2019. “Unemployed or disabled? Disability screening and labor market outcomes of youths.” University of Oslo Memorandum No. 05/2019, Department of Economics.


Chapter 1:

Pension reform and the efficiency-equity trade-off: Impacts of removing an early retirement subsidy
Pension reform and the efficiency-equity trade-off: Impacts of removing an early retirement subsidy

Asbjørn Goul Andersen a,⁎, Simen Markussen a, Knut Røed a

⁎ The Ragnar Frisch Centre for Economic Research, Oslo, Norway

A R T I C L E   I N F O

JEL codes:
H55
D31
J22
J26

Keywords:
Pension reform
Inequality
Labor supply

A B S T R A C T

We provide empirical evidence that the removal of work disincentives embedded in retirement earnings tests can increase old-age labor supply considerably, but it does so at the cost of more income inequality. To identify causal effects, we exploit a reform of the Norwegian early retirement program, which entailed that adjacent birth cohorts faced completely different work incentives from the age of 62. The reform removed a strict retirement earnings test such that pension wealth was redistributed from early to late retirees. Given pre-existing employment and earnings patterns, this implied a considerable rise in old-age income inequality. In theory, this direct increase in inequality could be either amplified or offset by changes in labor supply. We estimate that the reform triggered a 42% increase in average hours worked during the period covered by early retirement options; however, as labor supply responses were of similar magnitudes across the earnings distribution, they did little to modify the rise in inequality. As measured by the Gini coefficient, inequality in overall old-age income rose by approximately 0.03 (21%).

1. Introduction

In recent years, many developed countries have reformed their pension systems to address the rising fiscal costs of population aging. A key element in many of these reforms has been to encourage senior workers to postpone retirement. One strategy for promoting higher labor supply among the elderly is to remove the earnings test on pension income, such that workers above the threshold age for (early) retirement maintain strong incentives to work. This also removes an important source of economic inefficiency, as the retirement earnings test widens the wedge between employers’ wage costs and workers’ net pay considerably, discouraging work even when its social value by far exceeds the private value of the forgone leisure. However, the fact that not all workers have equal opportunities for extending their careers, e.g. due to poor health, outdated skills, or arduous work, has raised concerns about the distributional consequences of such policies.

Many countries still have earnings tests in various forms for individuals who retire before the statutory retirement age, including Germany, France, Belgium, Austria, and USA (see, e.g. OECD (2017, Table 2.A2.1) and Börsch-Supan et al. (2018, Table 1) for recent overviews). In the present paper, we examine to what extent removing a retirement earnings test (RET) and introducing actuarial neutrality in the pension system represent a tradeoff between equity and efficiency. We exploit a Norwegian pension reform implemented in 2011, which for a large group of workers transformed an earnings-tested early retirement program into an unconditional life-long pension annuity that could be claimed on actuarially neutral terms by every eligible worker from the age of 62, regardless of own labor earnings. The reform implied that pension entitlements previously reserved for those who actually left the labor market were distributed among all workers. As a result, the lifetime value of the new unconditional early retirement pension was reduced for workers who retired at the earliest possible occasion. For those who continued working into the early retirement period, the pension was increased.

Several studies have investigated the labor supply effects of policies relating to a retirement earnings test (RET). In general, the literature separates between two types of RETs, depending on deferral options. When deferral is possible on actuarially neutral terms, the earnings test is in some sense superficial, and, for a rational forward-looking agent, work incentives are largely unaffected. RET reforms of such schemes have been evaluated in both the US (Friedberg, 2000; Song and Manchester, 2007; Haidar and Loughran, 2008; Engelhardt and Kumar, 2009) and in the UK (Disney and Smith, 2002). To the extent that these studies find positive labor supply effects of the RET removal (e.g., Friedberg, 2000, and Engelhardt and Kumar, 2009), this is likely to reflect risk-aversion, short-sightedness, or simply failure to understand that withheld benefits are not lost, but just paid out later on (Brown et al., 2013; Rabinovich and Perez-Arce, 2019).

When deferral is not an option, the effect on work incentives is obvious: Any postponement of retirement reduces the lifetime pension en-
tirement. Baker and Benjamin (1999) evaluate a sequential elimination of such a “real” RET in Canada in the 1970s and estimate a 10 percentage points increase in full year work among 65–69 year olds. Brinch et al. (2017) use a difference-in-differences approach to study the effects of a stepwise real RET-removal in Norway during 2008–10 on the earnings of 67-year-old men. They find a sizeable positive earnings effect for workers who are still active at age 66. The pension reform examined in the present paper has also previously been evaluated in this context, disclosing a substantial overall labor supply effect (Brinch et al., 2015; Hernæs et al., 2016).

In summary, the existing empirical evidence suggests that abolishing (real) earnings tests on pension payments is an effective strategy for increasing labor supply among seniors. However, so far the distributional consequences of RET policies have received less attention. One notable exception is Bönke et al. (2018), who investigate the distributional effects of the introduction of an actuarial deferral option in the German early retirement system in 1992, which essentially removed a real RET. Their findings indicate large positive labor supply responses, at the cost of increased inequality. Another exception is Hernæs and Jia (2013), who investigate the distributional effects of a stepwise increase in the earnings threshold for RET in Norway in 2002 (applying at age 67–69). They find a positive labor supply effect at the intensive margin, driven by those who were still active at the age of 66 and had earnings around the thresholds. Since these thresholds were quite low, work incentives were primarily improved at low earnings, and, as a result, the reform led to a decrease in old-age earnings inequality. There is also a small related literature examining the distributional consequences of raising the early retirement age (Cribb and Emmerson, 2019; Morris, 2019; Geyer et al., 2020).

A priori, it is not clear how the labor supply responses to the RET removal affects the overall old-age labor earnings distribution. On the one hand, effects at the extensive margin should reduce overall income inequality, since richer people tend to work regardless of RET, and hence have less scope for increasing their labor supply. On the other hand, it has been argued that many elderly workers with physically demanding and poorly paid jobs do not really have the option of extending their career much beyond the early retirement age. These “worn-out workers” will thus become the losers in a regime where annual pensions are tightly attached to the age of actual retirement. Moreover, as pointed out by Egeton (2018), employees with low education and low pay are generally those who are most exposed to involuntary job loss and therefore have less possibilities to adjust the timing of retirement in accordance with own preferences.

Our empirical analysis builds on complete administrative data, covering the entire Norwegian population, with employer information and individual earnings trajectories from 1967 onwards. The data allow us to single out the group of private sector workers that was exposed to the removal of the earnings test (approximately 23% of the active workforce). Our primary empirical strategy is to compare the last two birth cohorts (1946–47) that were subjected to a real retirement earnings test with the first two cohorts (1949–50) that were exposed to a fully actuarially neutral pension system with no earnings test. The data allow us to compute virtually complete lifetime earnings histories for all these cohorts. We show that while the distribution of prime-age earnings – defined as average annual earnings over the 40-year period from age 21 to 60 – is almost identical for the pre- and post-reform cohorts, their earnings paths after the early retirement age (62 years) diverge considerably. Our analysis confirms the findings of Hernæs et al. (2016) of large average labor supply effects at age 63 and 64, and we are able to show that these effects remain strong at ages 65–67, and even stretch beyond statutory retirement to age 68 at which point work incentives were unaffected by the reform.

We carry out a novel empirical analysis in three parts. First, we explore how the labor supply responses vary across the prime-age earnings distribution. Our main strategy is to divide the sample into deciles based on accumulated labor earnings from age 21 to 60, and estimate the effect of the pension reform separately within each bin. We find that the labor supply responses to strengthened work incentives are surprisingly similar across the distribution of prime-age labor earnings. For all earnings deciles, except at the very top, employment rates during age 63–65 increased by approximately 20 percentage points, whereas (unconditional) hours worked per week increased by 7–10 h. During age 66–67, the employment rate increased by 10–15 percentage points and hours worked per week by 3–5 h. In total, we estimate that the reform caused an increase in hours worked by as much as 42% during the five-year early retirement period. Some of these effects remained even after the end of this period, despite almost unchanged economic incentives at this point. At age 68, we estimate a 4 percentage point increase in employment and a 2 h increase in work per week. In terms of employment status and hours worked, the weakest response is found among the top-earners, who had relatively high employment rates even prior to the reform and thus had less potential for an increase. In terms of absolute earnings, on the other hand, the effects are largest at the top of the prime-age earnings distribution.

Second, we characterize the winners and losers. As the reform essentially shifted pension wealth from early to late retirees, the clearest winners are those who would have preferred to continue working through the early retirement period in both regimes (the “always-workers”). For this group, the new pension entitlements can almost be considered an annual lump-sum transfer. The clearest losers are those who would have preferred to leave the labor market at the earliest possible occasion in both regimes (the “never-workers”). For these workers, the reform merely reduced the lifetime value of their early retirement pension (by approximately 21%). Assuming that nobody decides to leave (remain in) employment as a result of higher (lower) take-home wages, we can identify the definite winners of the pension reform as those who continued working until the statutory retirement age in the pre-reform period. Likewise, we can identify the definite losers as those who left the labor market at the lowest early retirement age in the post-reform regime. Defined this way, we find that 15% of the eligible workers can be counted as definite winners, whereas 6–7% is definite losers. Comparing these two groups, we show that the “always-working” winners tend to be individuals with higher prime-age earnings, higher education, more prestigious occupations, and much lower sickness absence in the past than the “never-working” losers.

Finally, we examine the distributional consequences of the reform more directly by examining its effect on the distribution of accumulated pension and labor income after age 62. In order to do so, we use the pre-reform cohorts to construct a sample that matches the post-reform cohorts on gender, prime-age earnings, and age 60 earnings, and treat the observed old-age outcomes for this adjusted sample as counterfactual observations for the post-reform sample. The resultant trajectories allow us to disentangle the effect of the new entitlement rules – given the pre-existing labor supply behavior – from the consequences of the reform-generated changes in labor supply. Our findings show that while roughly 40% of the workers lost out in terms of lower pension entitlements, the large labor supply responses ensured that the vast majority (93%) came out with higher overall old-age income. The new entitlement rules also led to a considerable increase in old-age income inequality, whereas the labor supply responses were more or less neutral in distributional terms. The resultant increase in income inequality turned out to be considerable. Measured by the Gini coefficient, overall old-age income inequality increased by approximately 21% as a direct result of the reform. However, in contrast to recent studies examining the impact of higher early retirement age (Cribb and Emmerson, 2019; Morris, 2019), we find no indications of severe poverty. This must be interpreted in light of the relatively resourceful group of workers that are included in our analysis. By focusing on private sector workers who are eligible for the early retirement program both before and after the reform, we essentially limit our attention to workers who are still in employment at age 60 and who have had stable careers over many years with relatively high earnings. Although this group of workers is not rep-
resentative for the full population of potential retirees in Norway, they may be quite representative for the type of workers that governments might successfully convince to extend their labor market careers.

2. Institutional setting: the Norwegian pension reform

The Norwegian pension system has three main pillars: (i) a universal public old-age pension from the National Insurance Scheme (NIS), (ii) contractual early retirement schemes (“Antalle-festet Pensjon” henceforth referred to by the acronym AFP), and (iii) occupational pension schemes in the public and private sector. The reform in 2011 entailed a major restructuring of the universal public pension system, introducing a tighter relationship between individual lifetime earnings and pension entitlements, longevity-adjusted annual pensions, and less generous indexation. However, these changes are implemented gradually and thus had very limited impact on the cohorts retiring around the time of the reform. Their longer-term distributional impact is evaluated in Nicolajsen and Stolen (2016) and Halvorsen and West Pedersen (2019).

In the present paper, we focus on a reform element that had large and immediate consequences for a large group of workers; namely the removal of the retirement earnings test for private sector workers qualifying for early retirement (AFP). This reform was implemented in a quasi-experimental fashion, in the sense that adjacent birth cohorts suddenly faced completely different early retirement incentives.

Prior to the reform, the AFP-scheme essentially offered a full pension from the age of 62 until the statutory retirement age of 67, when the old-age pension could be claimed. While it was possible to combine the AFP-pension and labor income, a confiscatory earnings test implied that the effective tax rates on continued work were very high; see Hernaes et al., (2016). There was no deferral option, so postponing retirement would reduce lifetime pension wealth. Moreover, full retirement at age 62 had no consequences for future pension entitlements, which were calculated as if the retiree had continued working as before until age 67. Workers therefore faced substantial disincentives to work after the age of 62.

For private sector workers, two elements of the reform greatly changed this; namely: i) the introduction of flexible take-up of the old-age pension from age 62 with no earnings test and with actuarially neutral adjustments of the pension; and ii) the restructuring of the AFP-scheme into a lifelong annuity, also available from age 62 with no earnings test and with actuarial neutrality. The revised system thus implied a complete decoupling of decisions regarding labor supply and decisions regarding the timing of pension claiming.

The new AFP applied to individuals who had not yet claimed AFP by January 2011, implying that the cohort of 1949 was the first to be fully covered by the new scheme. Individuals born in 1948 could choose to enroll in the new scheme by postponing take-up until 2011. This cohort will therefore consist of individuals enrolled in both the old and the new scheme. Individuals born in 1947, 1946, 1945, and 1944, who had still not taken-up AFP by January 2011, could also enroll; however, they were offered substantially less generous versions of the scheme (corresponding to 60%, 40%, 20%, and 10% of the full entitlement, respectively). In the following, we shall generally refer to the cohorts born in 1949 or later as the post-reform cohorts, while we refer to the cohorts born in 1947 or before as the pre-reform cohorts.

The restructuring of the private sector AFP-scheme was the result of tripartite negotiations between the state and the major associations of employers and employees, starting in 2008. In order to secure an agreement, the government provided extra funding, facilitating an extra “compensation benefit” for all workers born before 1963. Hence, as we show below, the majority (approximately 60%) of the workers came out with higher pensions than under the pre-reform regime. From a fiscal point of view, this turned out to be a good investment, though, as the extra tax revenue generated by the resultant labor supply responses more than compensated for the extra funding; see Hernaes et al. (2016). The outcomes of the AFP-negotiations and the main features of the new private sector AFP were probably known by most workers from around mid-2009. At this time, it was generally not possible to enroll into or switch between the schemes, since AFP-eligibility in both the private and most of the public sector requires several years of employer- and sector-specific tenure.

3. Data and identification strategy

Our empirical analysis exploits Norwegian administrative data containing detailed information on earnings, employment, education, attainment, pension entitlements, and demographic characteristics for the entire population. The main analyses are based on the birth cohorts who reached the age of 62 just before (born 1946–47) and just after (1949–50) the implementation of the reform. To assess pre-reform trends, we also include older cohorts (1943–45) in parts of the analysis. We exclude the 1948-cohort from the main part of the analysis because members of this cohort could self-select into either the old or the new AFP-scheme. We return to this cohort in Section 5, however, where we use it to identify the workers’ own preferences with respect to the choice of early retirement scheme.

Based on the entire earnings history from 1967 and information about the main employer in the years preceding the reform, we identify the old-age pension and AFP-entitlements at an individual level. The eligibility requirements for a full pension with AFP changed slightly as part of the reform; hence, to avoid selectivity, we restrict our sample to workers who would have qualified by age 62 under both the old and the new rules (see Online Appendix A for a description of eligibility rules before and after the reform). In order to minimize potential endogeneity problems related to anticipation of the reform, we condition our sample on employment and eligibility by age 60 rather than by age 61 or 62 (because incentives to stay on until age 61 or 62 may have been affected by the reform). Descriptive statistics for the pre- and post-reform cohorts are presented in Table 1 and Fig. 1. We note that the two groups are similar in terms gender, fraction of immigrants, educational attainment, work hours, and earnings. The latter is particularly evident when we look at the distribution of prime-age earnings (average annual earnings from age 21 to 60) for the pre- and post-reform cohorts. As can be seen from Fig. 1, panels (a) and (b), the distribution functions for pre- and post-reform cohorts are hardly distinguishable. The earnings levels observed at age 60 and 61 are somewhat lower for the post-reform cohorts, however, most likely because these cohorts were adversely affected at this age by the economic downturn in 2009–2010 following from the financial crisis.

The main outcome variables used in the analyses are employment status, gross (pre-tax) earnings, and weekly work hours in the calendar years at which the individuals reach the age of 63, 64, 65, 66, and 67. Data on earnings come from the public tax records, and individuals with annual earnings exceeding NOK 100,000 (in 2020 value, corresponding to € 10,000 or $ 11,000) are classified as employed. This threshold is probably too high in the early years before the economic downturn.

Since the negotiations of the reform began in 2008, and the youngest post-reform cohort reached the age of 60 in 2010, we cannot completely rule out behavioral responses to the reform before age 60. As a robustness check, Hernaes et al. (2016) carry out their analyses conditioning on employment at age 58. The fact that this does not noticeably change their results indicates that age selection seems to be a minor concern.

Given that reliable earnings data are available for whole calendar years only, the outcomes used in this paper are also defined at the calendar year level. We start with the year individuals reach the age of 63 (and thus are 62 years old at the start of the year), since this is the first year where we can observe the full effect of the reform.

Earnings obtained in other years are inflated to 2020 value using the adjustment factor in the Norwegian social insurance system, which corresponds approximately to the annual average wage growth.

---

1 The pension was reduced in proportion to the income as a share of previous income (defined as the average income in the three best of the last five years).
implies that a person is considered employed in a given year if annual earnings exceeded approximately 18% of the average earnings level for a full-time-full-year position. Weekly work hours are calculated using an hourly wage rate imputed from earnings and work hours at age 60.

It is clear from Fig. 1 that while the distribution of cumulative labor earnings up to age 60 are virtually identical for the pre- and post-reform cohorts (panels (a) and (b)), their earnings after age 62 diverge considerably (panels (c) and (d)). In particular, we note a large drop in the spike at zero earnings and an increase in the probability mass around typical full-time earnings (panel (c)), implying that the old-age cumulative earnings distribution (panel (d)) is significantly shifted to the right for the post-reform cohorts.

Given the striking similarity of the pre- and post-reform cohorts' earnings paths up to age 60, we base identification of the reform effects on a direct comparison of these cohorts' employment and earnings patterns from age 63 onwards (i.e., from the age at which the reform had a full effect), with controls for observed individual characteristics. The main identifying assumption underlying our empirical strategy is that the two last pre-reform cohorts represent a valid counterfactual for the two first post-reform cohorts. In other words, we assume that if the reform had never been enacted, the labor supply behavior (and outcomes) of the post-reform cohorts would have been largely identical to that of the pre-reform cohorts (after controlling for observable differences between the groups). This translates into three different assumptions, discussed in turn below, namely: (i) no self-selection into or out of the analysis population, (ii) no calendar time effects, and (iii) no spillovers between members of the pre- and post-reform cohorts, implying satisfaction of the so-called Stable Unit Treatment Value Assumption (SUTVA).

As discussed in Section 2, self-selection related to anticipation of the reform cannot be entirely ruled out. While selection into the private sector AFP-scheme was generally not possible, selection out of the scheme and into the public sector scheme may have been an option for some. If post-reform workers, who wish to retire at an early stage, were more likely to shift to the public sector, we might overestimate the true reform effect, because the remaining members of the post-reform group are more prone to continue working. The fact that we condition the sample on employment and AFP-affiliation at age 60 leaves little room for such a response, however, since the post-reform cohorts reached this age in 2009 and 2010, respectively, shortly after the content of the reform was known. Hernæs (2017) shows that less than half a percent of private sector workers eligible for the post-reform AFP switches to the public sector between age 59 and age 60. Moreover, Hernæs et al. (2016) find that conditioning the sample on employment at age 58 instead, does not alter the estimated labor supply responses noticeably, but does introduce more noise due to a less accurate determination of AFP-eligibility. This indicates that endogeneity in the AFP-entitlement is unlikely to be driving any of the results.

To assess the validity of the assumption of no calendar time effects, either related to underlying trends or to cyclical fluctuations, panels (a) and (b) of Fig. 2 show age-specific employment rates and average annual earnings for individuals qualifying for the early retirement scheme. The statistics are shown for the last five pre-reform birth cohorts; i.e., those born in 1943, 1944, 1945, 1946, and 1947, and for the first two post-reform cohort (1949 and 1950). Focusing on the labor supply at age 63–64, there are no indications of a trend toward increased labor supply among the pre-reform cohorts. It is perhaps possible to see a slight trend toward higher employment rates at age 65–66, but that could be related to the fact that the latest pre-reform cohorts were partially treated at this point, provided that they had not already enrolled into the old AFP; conf. Section 2. In any case, the main take-away from Fig. 2 is that the big shifts coincided with the reform. For comparison, panels (c) and (d) of the figure shows the age-specific employment rates and earnings for a group of workers that were not affected by the pension reform.
We observe very small changes in employment and earnings for this group.\(^5\)

It is also worth noting that the outcome period used in our analysis was a period of relative macroeconomic stability, particularly during the first four years (2009–2013) when the unemployment rate fluctuated between 3% and 4%. After that, the economy lost some steam, and the unemployment rate peaked around 5% in 2015. If anything, this development should have contributed to lower lowering in the post-reform cohorts during the ages covered by early retirement options. Fig. 3 below shows the employment rates for the different age groups of the entire population around the time of the reform. We see that the employment rates are relatively stable for 50–54 year-olds, 55–59-year olds, and 60–62 year old, whereas it jumps discretely at age 63 after the reform.

Spillover effects between birth cohorts cannot be entirely ruled out. On the one hand, increased labor supply of the post-reform cohorts at the age of 62 and 63 could harm the employment prospects of pre-reform individuals at the age of 65 and 66, who might be competing for the same kinds of jobs. However, only a small minority of workers will be competing for new jobs at this age, whereas the grand majority either remain in their current job (perhaps working fewer hours) or fully retire. This type of spillover effects should therefore be negligible. Another kind of spillover could arise from the joint retirement decisions of married couples. Kruse, (2021) provides empirical evidence from Norway suggesting that spousal spillovers in retirement decisions are asymmetric, such that wives respond to their husbands’ choices, but not necessarily vice versa. Given the typical age difference within couples, this implies that the most relevant spillover effect in our data is a situation where a male worker belonging to the pre-reform cohort chooses to retire early due to the poor work incentives, and that this instigates his younger wife, belonging to a post-reform cohort, to retire as well. This implies that the full reform effects will not be revealed until both spouses have entered the post-reform regime. For our analysis, it implies that the ultimate reform effects might be somewhat underestimated.

\(^5\) Only workers that were neither part of the AFP agreement nor eligible for the new early retirement option in the public pension system were unaffected by the pension reform. This group makes up approximately 11% of the workforce and consists of workers with relatively low and/or unstable previous earnings.
Fig. 3. Employment rates for the entire population in Norway. By age and year. Note: Employment is defined as having annual earnings above NOK 100,000 (measured in 2020-value).

Table 1
Descriptive statistics.

<table>
<thead>
<tr>
<th></th>
<th>Pre reform cohorts (Born 1946–47)</th>
<th>Post reform cohorts (Born 1949–50)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of individuals</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>16,109</td>
<td>15,628</td>
</tr>
<tr>
<td><strong>Share of all employed at age 60 (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>23.5</td>
<td>23.6</td>
</tr>
<tr>
<td><strong>Baseline characteristics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Women (%)</td>
<td>19.1</td>
<td>21.6</td>
</tr>
<tr>
<td>Immigrants (%)</td>
<td>0.7</td>
<td>1.1</td>
</tr>
<tr>
<td>Compulsory education only (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>18.7</td>
<td>17.0</td>
</tr>
<tr>
<td>High school (%)</td>
<td>62.4</td>
<td>64.8</td>
</tr>
<tr>
<td>College (%)</td>
<td>18.9</td>
<td>18.2</td>
</tr>
<tr>
<td>Weekly work hours at age 60</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>41.3</td>
<td>41.0</td>
</tr>
<tr>
<td>Months of sick leave last 15 years (annualized)</td>
<td>0.36</td>
<td>0.39</td>
</tr>
</tbody>
</table>

| **Earnings (NOK 1000):** |                                   |                                   |
| at ages 21–60 (annualized) | 612.6                             | 614.7                             |
| at age 60                  | 752.9                             | 724.6                             |
| at age 61                  | 720.2                             | 688.7                             |
| at age 62                  | 609.9                             | 625.1                             |
| at age 63                  | 418.5                             | 518.2                             |
| at age 64                  | 323.0                             | 449.8                             |
| at age 65                  | 264.1                             | 376.3                             |
| at age 66                  | 221.8                             | 293.3                             |
| at age 67                  | 167.8                             | 205.5                             |
| at age 68                  | 102.8                             | 129.1                             |

**Characteristics of occupation at age 60:**
- Life expectancy at age 62 (years): 21.5, 21.6
- Social class (ISIE scale): 47.2, 47.1

Note: All earnings are measured in NOK 1000 and inflated to 2020-value (using the deflator in the Norwegian pension system). Sick leave in the last 15 years (before age 60) is calculated as the average number of months per year with any registered sick leave (only sick-leave periods exceeding 16 days in duration are registered). Life expectancy at age 62 is based on occupation- and gender-specific estimates of Borgan and Tixmon (2015). Social class refers to the status of the occupation held at age 60 according to the International Socio-Economic Index of occupational status (ISEI) suggested by Ganeboom et al. (1992), which is based on the International Standard Classification of Occupations (ISCO).

4. The social gradient in labor supply responses

In order to assess the potential heterogeneity in reform effects, we divide the population into different socioeconomic groups based on information available at age 60, and estimate separate reform effects for each group. Given our focus on the distributional consequences of the reform, we use prime-age earnings as the primary grouping criterion; i.e., we divide the population of workers at age 60 into deciles based on each worker’s position in the age 21–60 earnings distribution within own birth cohort. Fig. 4, panel (a), presents the average age 21–60 earnings levels for each of these deciles, measured in 1000 NOK (inflated to 2020-value), for all the four birth cohorts included in our estimation sample. Average earnings over these 40 years vary from around 325,000 NOK in the lowest decile to more than one million NOK in the upper decile. Panel (b) then shows, for each decile, the impact of the RET reform on the economic reward (net of tax) associated with continuing another year (at age 63) with the job held at age 60, while panel (c) shows the relative increase in this reward. It is clear that the improvement in work incentives is very large across the earnings distribution, with the average annualized improvement varying between NOK 175,000 and 230,000 measured in absolute terms and between 50 and 200 percent measured in terms of relative improvement. While the absolute increase in the take-home wage was largest at the top of the earnings distribution, the relative increase was largest at the bottom.

Fig. 5 illustrates how a classification of workers based on prime-age earnings correlates with a range of individual characteristics. Panel (a) first shows how the prime-age earnings levels at age 60 vary across the deciles in the accumulated prime-age earnings distribution. A first point to note is that the earnings levels are relatively high at this age for all the deciles in our estimation sample, reflecting that we have conditioned on employment and early retirement eligibility. For the sample as a whole, the observed average earnings level at age 60 of around NOK 740,000 lies around 35% above the average full-time-full-year earnings observed for all workers in Norway. Yet, the earnings differences are substantial, with the top decile earning approximately three times as much as the bottom decile. Panel (b) then illustrates the large gender gap in prime-age earnings within these birth cohorts. While women constitute 20% of the whole sample, they make up as much as 80% of the bottom decile and as little as 1% of the top decile. Panels (c)–(f) show how a range of alternative classification indicators differ across the prime-age earnings deciles; i.e., educational attainment (panel (c)), the social status of the occupation held at age 60 (panel (d)), the expected longevity associated with the occupation held at age 60 (panel (e)), and overall sickness absence during age 45–60 (panel (f)). It is evident that the categorization based on prime-age earnings correlates closely with alternative categorizations based on these characteristics. We return to estimates based on such alternative categorizations after we have presented the main results.

The estimation of group-specific reform effects is based on a simple ordinary least squares regression of the following type:

\[ Y_{it} = \beta_0 x_i + \beta_1 T_{i} + \epsilon_{it}, \]

where \( Y_{it} \) represents the outcome of interest (employment, earnings, weekly work hours) for a person \( i \) belonging to prime-age earnings decile \( d \) measured at age \( t \), and \( x_i \) is a vector of covariates including gender, education (nine fields and eight levels), country of origin for immigrants (five regions), and weekly work hours and earnings at age 60. \( T_i \) is a treatment-dummy equal to 1 for the post-reform cohorts, and 0 for the pre-reform cohorts, and the coefficient \( \beta_1 \) represents the treatment effect. This equation is identical to the one used for the whole population in Hernes et al. (2016), and, for ease of comparison, we also use exactly the same explanatory variables. Note, however, that we use a more restric-
tive definition of employment, as we require annual earnings to exceed NOK 100,000 (rather than 10,000). Our definition still allows for relatively minor positions, given that NOK 100,000 constitutes less than a fifth of the average earnings level for a full-time position in Norway.

Figs. 6–9 present our main results, in terms of estimated effects of the reform on employment status, weekly hours of work, annual labor earnings, and annual labor earnings relative to the earnings level at age 60, respectively. Starting with employment status, the top panels of Fig. 6 show the employment rates at age 63, 64, 65, 66, and 67, respectively, within each prime-age (21–60) earnings decile for the pre-reform and post-reform cohorts. We see that the employment rate increases along the distribution of past earnings for both groups. The differences in employment levels between the pre- and post-reform cohorts appear to be roughly constant across the earnings distribution. The bottom panels report the reform effects on employment estimated within each decile with a 95% confidence interval. The effects estimated for the whole sample (indicated by the dashed horizontal line) were roughly 17, 22, 21, 16, and 10 percentage points at age 63, 64, 65, 66, and 67, respectively. The within-decile estimates are generally around the same level across the earnings distribution, with a moderate hump-shape at ages above 63 such that the effects are largest at the upper-medium part of the distribution, but smallest at the very top. This pattern repeats itself also for the hours worked outcome; see Fig. 7. At age 63, weekly hours worked increased by approximately 7 h throughout the earnings distribution. At higher ages, a more conspicuous hump shape emerges, with the largest effects at the upper-medium part of the distribution and lower effects at the top.

Although the estimated reform effects on employment and hours worked are roughly the same across the prime-age earnings distribution, measured in absolute terms, it is worth noting that relative to the initial (pre-reform) level of labor supply, the effects are considerably larger at the bottom of the earnings distribution. For example, while the seven added work hours supplied at age 63 by individuals in the bottom of the prime-age earnings distribution constitutes a 35% increase relative to pre-reform hours, the same number of added hours toward the upper part of the distribution constitutes a 25% increase. Considering the reform effects for all years (age 63–67) together, we estimate that weekly hours worked increased by 6.1 on average, or by 42%. For the bottom decile, it increased by 5.4 h (42.6%). The effect reached its maximum for the 7th decile with 7.3 h (51%), and its minimum for the very top decile with 4.3 h (21.9%).

The estimated reform effects on annual earnings are provided in Fig. 8. On average, labor earnings increased by 100–150,000 NOK in each year with entitlement to early retirement. For this outcome, there is a marked positive social gradient in the effect pattern, with larger reform effects the higher the position in the prime-age earnings distribution. Note that it is not meaningful to estimate the earnings effects with the conventional log-specification in our case, as the behavioral responses primarily occur at the extensive margin, with earnings typically either equal (or close) to zero or equal (or close) to the age 60 level; see Fig. 1, panel (c). A more appropriate alternative may be to define the outcome explicitly in terms of earnings relative to the age 60 level. The results from such a model are presented in Fig. 9. The effects are again very similar across the earnings distribution, and conspicuously similar to the employment effects shown in Fig. 6. At ages 63–65, the effects on annual earnings constitute approximately 15–20% of the initial (age 60) earnings level for all deciles in the earnings distribution, except for the top decile, where the effects again are significantly smaller than for the other groups.

Figs. 6–9 show estimation results for the five-year period that best matches the early retirement period in which work incentives changed (ages 63–67). Given the large labor supply effects identified until the statutory retirement age at age 67, one could hypothesize spillover effects also into higher ages. In Appendix B, we show estimation results also for age 68, indicating positive reform effects on employment (4 percentage point) and work hours (2 h per week) even at this stage. For one of the post-reform cohorts, we also observe outcomes at age 69, but we do not find any significant effects at this age (not shown).

The extensive nature of labor supply decisions made during the early retirement age makes it impossible to estimate meaningful labor supply elasticities at the individual level. However, a natural way to sum up the messages from Figs. 6–9 could be to compute such elasticities at the group-level; i.e., divide the decile-specific reform-initiated relative changes in earnings or hours worked by the corresponding relative changes in take-home wages reported in Fig. 4, panel (c). If we

---

6 In Appendix B, we also present results for age 62. As the outcome year corresponding to each age is defined based on the calendar year in which the indicated age is reached, some (but not full) effects could be expected already at this age. The results shown in the appendix indicate a 2-3 hours increase in weekly work, but no impact on employment based on the employment definition used in this paper.
Fig. 5. **Individual characteristics by decile in the prime-age (21–60) earnings distribution.** Note: The reported statistics are based on the total estimation sample, consisting of AFP-eligible workers belonging to the 1946, 1947, 1949, and 1950 birth cohorts (\(N = 31,738\) except in panels (d) and (e) where missing information on occupational classification reduces the sample to \(N = 31,021\) and \(N = 26,211\), respectively). See the note to Table 1 for a description of how we have defined and computed social class, life expectancy, and sick leave.

Fig. 6. **Observed employment rates for pre- and post-reform cohorts and estimated reform effects by age and decile in the age 21–60 earnings distribution.** Note: The top diagrams indicate the employment rate at age 63–67 across the earnings distribution for the pre-reform cohorts (1946–47, gray dots) and post-reform cohorts (1949–50, black dots), respectively. Earnings deciles are based on earnings at age 21–60 and are calculated within cohorts. The lower diagrams report the estimated reform effects (with 95% confidence intervals) for each decile, based on Eq. (1), as well as the average effect size across the income distribution (the dashed horizontal line). The population consists of workers affiliated with a private sector AFP scheme who were employed at age 60 and qualified for retirement at age 62 both before and after the reform.
do this, we obtain elasticity estimates that apparently rise monotonically with prime-age earnings, from 0.2 to 0.3 for the lowest deciles to 0.3–0.5 for the upper deciles. However, this would arguably give a distorted picture of group-specific labor supply responses. As noted by Hernes et al. (2016), given that there was a strictly positive labor supply within all groups even before the reform, it is difficult to imagine anything more than a doubling of the labor supply for this group (in which case absolutely everyone works full time). Since we know from Fig. 4 that the take-home wage was more than tripled for this group due to the reform, this imposes an absolute upper limit on the labor supply elasticity calculated this way of approximately 0.5. By contrast, the top decile would reach full-time work for everyone with a 50% increase in labor supply; hence, given that their take-home wage also increased by 50% on average (Fig. 4), the absolute upper limit on their elasticity calculated this way is approximately 1.0.

Viewed as a whole, we interpret the results in Figs. 6–9 as suggestive of relatively homogenous labor supply responses across the different earnings groups, with a possible exception for those with the highest earnings. This is somewhat surprising, since we would generally expect that “worn-out workers”, holding the most physically demanding jobs in the lower end of the earnings distribution, should respond less, having less scope for individual adjustments. One explanation may be that there is quite some overlap in occupational groups between deciles, such that low-wage individuals with long careers may fall into the same category as high-wage individuals with shorter or interrupted careers. This point suggests that it may be of some interest to assess alternative categorizations of socioeconomic groups. Hence, as an alternative to deciles based on accumulated prime-age earnings, we have divided the population into cells based on the occupation held by age 60. Fig. 10 presents

---

Fig. 7. Observed weekly hours worked for pre- and post-reform cohorts and estimated reform effects by age and decile in the age 21–60 earnings distribution. Note: The top diagrams indicate average hours worked at age 63–67 across the earnings distribution for the pre-reform cohorts (1946–47, gray dots) and post-reform cohorts (1949–50, black dots), respectively. Earnings deciles are based on earnings at age 21–60 and are calculated within cohorts. The lower diagrams report the estimated reform effects (with 95% confidence intervals) for each decile, based on Eq. (1), as well as the average effect size across the income distribution (the dashed horizontal line). The population consists of workers affiliated with a private sector APP scheme who were employed at age 60 and qualified for retirement at age 62 both before and after the reform.

Fig. 8. Observed annual earnings for pre- and post-reform cohorts and estimated reform effects by age and decile in the age 21–60 earnings distribution. Note: The top diagrams indicate average earnings at age 63–67 across the age 21–60 earnings distribution for the pre-reform cohorts (1946–47, gray dots) and post-reform cohorts (1949–50, black dots), respectively. The lower diagrams report the estimated reform effects (with 95% confidence intervals) for each decile, based on Eq. (1), as well as the average effect size across the income distribution (the dashed horizontal line). The population consists of workers affiliated with a private sector APP scheme who were employed at age 60 and qualified for retirement at age 62 both before and after the reform.
the result from this exercise. To facilitate comparison across the different categorizations, we show the average estimated effects for ages 63–67 combined instead of separate effects for each age. The first column of panels in Fig. 10 summarizes the effects already presented in Figs. 6–9, by reporting the estimated effects on average annual earnings during the full early retirement period. The two next columns then present corresponding effects by deciles in distributions based on occupation. In the second column (panels b), (f), (j), and (n)), the deciles are based on the occupations’ socioeconomic status according to the ISEI index (Ganzeboom et al., 1992), whereas in the third column (panels c), (g), (k), and (o)), they are based on occupation-by-gender-specific life expectancies (Borgan and Tjønneland, 2015). Finally, the last column in Fig. 10 (panels d), (h), (l), and (p)) presents results by decile in the distribution of accumulated sick-leave days over the past 15 years, sorted from those with most to those with least absence (as approximately 30% of the workers had zero absence, the rightmost data-point comprises more observations than the others). It seems clear that the labor supply responses are similar across the different socioeconomic groups regardless of the specific variable used to construct them. In particular, it is worth noting that labor supply sensitivity is almost unrelated to past sickness absence.

The choice of socioeconomic indicator has a large influence on the gender-composition of the various deciles. This is illustrated in the four lower panels of Fig. 10. We already know from Fig. 5 (panel b)) that based on accumulated prime-age earnings, we obtain a distribution heavily dominated by women at the lower end of the distribution and even more dominated by men at the top. Using the occupation-by-gender-specific life-expectancy measure, we get exactly the opposite pattern. This appears to have remarkably little influence on the distribution of estimated effects, however, suggesting that men and women respond similarly to work incentives. This is indeed confirmed by gender-specific estimates, which we report in Online Appendix C.

5. Characterization of winners and losers

The reform created winners and losers. Those who would have been fully employed under both regimes (“always-workers”) simply got a top-up pension from the new AFP-scheme as a bonus, while those who would have retired completely regardless of regime (“never-workers”) experienced a reduction in lifetime pension entitlements. Individuals who would have retired later in the new than in the old regime (“compliers”) could be better or worse off than before. We do not observe the compliers in the data, but if we impose a monotonicity assumption – i.e. assume that the reform had a weakly positive effect on labor supply for everyone – we are able to identify the always-workers in the pre-reform cohorts and the never-workers in the post-reform cohorts. We can think of the always-workers as those who were fully employed throughout the early retirement period in the pre-reform cohorts, despite the strong incentives to retire, and the never-workers as those who retired completely at age 62 in the post-reform cohorts, despite the strong incentives to continue working.

Table 2 reports the characteristics of always-workers and never-workers. By comparing the characteristics of these two groups, we can assess the composition of definite winners and losers. A first point to note is that there are more definite winners (15.0%) than there are definite losers (6.5%). Moreover, the group of winners consists of people with better education, higher prime-age earnings, more prestigious occupations, higher life expectancy, and less past sick leave than the group of losers. The differences in prime-age earnings appear to be moderate (6% higher in the winner group). For some of the other characteristics, the differences are considerable. For example, the losers have had roughly 2.5 times more sick leave than the winners have during the last 15 years. In addition, the occupational status codes suggest that winners are much more likely to have high-status occupations than losers are. The most heavily overrepresented occupations among the always-workers turn out to be machine— and plant operators, whereas the most overrepresented occupations in the always-worker group are architects, engineers, and managers (not shown in the table).

Fig. 11 shows how the fractions of never-workers (definite losers) and always-workers (definite winners) by decile in the prime-age earnings distribution. With notable exceptions for the bottom and top deciles, the fraction of never-workers appears to decline monotonically with prime-age earnings rank. The fraction of always-workers is relatively stable through the bottom half of the prime-age earnings distribution, and then rises steeply with earnings through the upper half, again with the extreme top as an exception.

Another way of assessing the distribution of winners and losers is by studying the behavior of the members of the 1948-cohort who could choose between the old and the new AFP. As explained in Section 2, the reform was implemented such that the enrollment into the old AFP had to be done before January 1, 2011. Given that enrollment was possible from the month after reaching the age of 62, workers born in November

Fig. 9. Observed annual earnings relative to earnings at age 60 and estimated reform effects by age and decile in the age 21–60 earnings distribution. Note: The top diagrams indicate average earnings, measured relative to earnings at age 60, at age 63–67 across the age 21–60 earnings distribution for the pre-reform cohorts (1946–47, gray dots) and post-reform cohorts (1949–50, black dots), respectively. The lower diagrams report the estimated reform effects (with 95% confidence intervals) for each decile, based on Eq. (1), as well as the average effect size across the income distribution (the dashed horizontal line). The population consists of workers affiliated with a private sector AFP scheme who were employed at age 60 and qualified for retirement at age 62 both before and after the reform.
1948 could choose almost freely between the old and the new scheme. In order to be part of the old system, they would have to take up the pension immediately after reaching age 62, whereas postponing take-up by a month (or more) would entail enrollment in the new system. Workers born earlier in 1948 could also choose between the two schemes, but would have to postpone take-up for 2–11 months, depending on month of birth, in order to enroll in the new scheme and avoid the early retirement earnings test. Fig. 12 shows the fractions of workers who actually chose the old AFP within these two populations, by decile in the prime-age earnings distribution. It is clear that the old AFP was more popular among workers in the lower end of the prime-age earnings distribution. Among workers born in November (Fig. 12, panel (b)), approximately 30% revealed a preference for the old earnings-tested AFP. However, while the fraction preferring the old AFP in the bottom decile of the prime-age earnings distribution was approximately 40%, it was just 10% in the top decile. Hence, there is a strong social gradient in the valuation of the reformed scheme.

6. Consequences for the old-age income distribution

To shed further light on the distributional consequences of the reform, we now examine its overall impacts on old-age income inequality. We do this by matching each member of the post-reform cohort to a similar person in the pre-reform cohort, and then comparing the resultant pre- and post-reform old-age income distributions. More specifically, we employ 1-to-1 nearest neighbor matching (with replacement) consisting of two steps. First, we match exactly on gender and percentile in the prime-age earnings distribution. Among the several potential matches from the first step, we then select the one who is most similar in terms of earnings at the age of 60. We then treat the entire earnings trajec-
Table 2
Characteristics of definite winners (always-workers) and losers (never-workers).

<table>
<thead>
<tr>
<th></th>
<th>Whole sample</th>
<th>Always-workers (Winners) and Never-workers (Losers)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I</td>
<td>II</td>
</tr>
<tr>
<td>Number of individuals</td>
<td>16,109</td>
<td>15,628</td>
</tr>
<tr>
<td>Share of pre-/post- group (%)</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>Baseline individual characteristics:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Women (%)</td>
<td>19.1</td>
<td>21.6</td>
</tr>
<tr>
<td>Immigrants (%)</td>
<td>0.7</td>
<td>1.1</td>
</tr>
<tr>
<td>Compulsory education only (%)</td>
<td>18.7</td>
<td>17.0</td>
</tr>
<tr>
<td>High school (%)</td>
<td>62.4</td>
<td>64.8</td>
</tr>
<tr>
<td>College (%)</td>
<td>18.9</td>
<td>18.2</td>
</tr>
<tr>
<td>Weekly work hours at age 60</td>
<td>41.7</td>
<td>41.3</td>
</tr>
<tr>
<td>Months of sick leave last 15 years (annualized)</td>
<td>0.355</td>
<td>0.391</td>
</tr>
<tr>
<td>Earnings at age 21–60 (annualized)</td>
<td>612.7</td>
<td>614.7</td>
</tr>
</tbody>
</table>

Characteristics of occupation at age 60
Life expectancy, by gender (years from age 62)
- Men: 21.5
- Women: 21.6
Social class (ISEI-scale mean)
- Low-ISEI occupations (lower-quartile,%): 47.2
- Medium-ISEI occupations (mid-quartiles,%): 48.7
- High-ISEI occupations (upper-quartile,%): 24.1
Lifetime pension entitlements (age 63–83) under different rules (1000 NOK)
- Pre-reform AFP: 1090
- Pre-reform public old-age pension: 5156
- Post-reform overall annualized pension: 302
Change in overall annualized pension (%)
- Change: 3.8
- Change at age 66: 25.7
Old-age earnings (age 63–83)
- Earnings: 1520
- Earnings annualized: 71

Note: Columns I and II present the distributions of characteristics in the total samples of individuals belonging to the pre- and post-reform cohorts, respectively. Column III reports the distribution for individuals in the pre-reform period who continue working as before (at least 80 percent of earnings level at age 60) every year up to (and including) age 66 (i.e. the always-workers or winners). Column IV reports the distribution for individuals in the post-reform cohorts who do not work at all after age 62 (i.e. the never-workers or losers). All monetary amounts are measured in NOK 1000 and inflated/deflated to 2020-value. See the note to Table 1 for a description of how we have defined and computed social class, life expectancy, and sick leave.

Fig. 11. Fractions of never-workers and always-workers by decile in the prime-age earnings distribution. Note: The never-workers are individuals in the post-reform cohorts who do not work at all after age 62 and the always-workers are individuals in the pre-reform cohorts who continue working as before (at least 80 percent of earnings level at age 60) every year up to (and including) age 66. The solid lines are a second order regression lines (OLS) through the ten respective data-points.
tory of the match from age 60 as the counterfactual earnings trajectory. Earnings are observed up to, and including, age 68 for all the individuals in our dataset. Although we would expect some of the labor supply responses during the age 63–68 period to imply slightly higher post-reform employment also from age 69, we choose the more conservative assumption of a zero reform effect after age 68 here. This assumption is implemented by setting labor earnings for everyone to zero from age 69 and onwards. We also set the expected lifetime to 83 years for everyone. Pension entitlements up to this age are then accurately computed.7

We explore how the reform affected the overall distribution of expected annual old-age income by comparing three different situations:

i) Pre-reform pension system and pre-reform labor supply
ii) Post-reform pension system and pre-reform labor supply
iii) Post-reform pension system and post-reform labor supply

Fig. 13 first provides average old-age earnings plotted against average prime-age earnings for each decile in the prime-age earnings distribution.8 It is clear that average pension income remained stable or increased slightly across the prime-age earnings decile bins, and it increased more in the upper part of the distribution (panel (a)). Labor earnings increased considerably for all groups, and again they increased more the higher the prime-age earnings (panel (b)). As a result, the relationship between prime-age earnings and old-age income became steeper (panel (c)).

Fig. 14 provides a more complete picture of the old-age total income distribution, in the form of densities (panel (a)), cumulative distribution functions (panel (b)), and Lorenz curves (panel (c)). Disregarding labor supply responses, the new entitlement rules shifted probability mass toward the tails of the distribution, and, hence, increased the degree of dispersion. Without labor supply responses, approximately 40% of the workers would have lost and 60% would have gained in terms of pension entitlements. However, the labor supply responses shifted the income distribution considerably to the right, and the vast majority (approximately 93%) of the workers thus came out with higher old-age

---

Fig. 12. Fraction choosing the old rather than the new AFP in the 1948 birth cohort. By decile in the prime-age earnings distribution. Note: The solid lines are a second order regression lines (OLS) through the ten respective data-points.

curves hardly the The for redistribution reform), terms in the years. The A.G. 63–67), Gini capped for the pension administration, for primes (age 21–60) women (age 63–83) income: Old-age income: overall income: from the period (63–67) Old-age (63–83) Total lifetime (21–83) the cohort. Overall income: Old-age (63–83) Total lifetime (21–83) Men Earnings: Prime-age (21–60) 0.162 0.162 0.162 Overall income: Old-age (63–83) Total lifetime (21–83) Women Earnings: Prime-age (21–60) 0.151 0.151 0.151 Overall income: Old-age (63–83) Total lifetime (21–83) A. All Earnings: Prime-age (21–60) 0.181 0.181 0.181 Early retirement period (63–67) 0.581 0.581 0.464 Overall income: Old-age (63–83) 0.129 0.157 0.157 Total lifetime (21–83) 0.168 0.169 0.169 Table 3 Income inequality (Gini coefficients) before and after the pension reform.

<table>
<thead>
<tr>
<th></th>
<th>I Pre-reform pension rules and pre-reform labor supply</th>
<th>II Post-reform pension rules and pre-reform labor supply</th>
<th>III Post-reform pension rules and post-reform labor supply</th>
<th>IV Total reform effect (III-I)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. All</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Earnings:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prime-age (21–60)</td>
<td>0.181</td>
<td>0.181</td>
<td>0.181</td>
<td>-0.117 (20.1%)</td>
</tr>
<tr>
<td>Early retirement period (63–67)</td>
<td>0.581</td>
<td>0.581</td>
<td>0.464</td>
<td></td>
</tr>
<tr>
<td>Overall income:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Old-age (63–83)</td>
<td>0.129</td>
<td>0.157</td>
<td>0.157</td>
<td>0.028 (21.4%)</td>
</tr>
<tr>
<td>Total lifetime (21–83)</td>
<td>0.168</td>
<td>0.169</td>
<td>0.169</td>
<td>0.001 (0.6%)</td>
</tr>
<tr>
<td>B. Men</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Earnings:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prime-age (21–60)</td>
<td>0.162</td>
<td>0.162</td>
<td>0.162</td>
<td>-0.119 (-20.4%)</td>
</tr>
<tr>
<td>Early retirement period (63–67)</td>
<td>0.583</td>
<td>0.583</td>
<td>0.464</td>
<td></td>
</tr>
<tr>
<td>Overall income:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Old-age (63–83)</td>
<td>0.128</td>
<td>0.155</td>
<td>0.155</td>
<td>0.027 (21.1%)</td>
</tr>
<tr>
<td>Total lifetime (21–83)</td>
<td>0.153</td>
<td>0.155</td>
<td>0.155</td>
<td>0.002 (1.3%)</td>
</tr>
<tr>
<td>C. Women</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Earnings:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prime-age (21–60)</td>
<td>0.151</td>
<td>0.151</td>
<td>0.151</td>
<td>-0.108 (19.8%)</td>
</tr>
<tr>
<td>Early retirement period (63–67)</td>
<td>0.545</td>
<td>0.545</td>
<td>0.437</td>
<td></td>
</tr>
<tr>
<td>Overall income:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Old-age (63–83)</td>
<td>0.102</td>
<td>0.137</td>
<td>0.137</td>
<td>0.035 (34.3%)</td>
</tr>
<tr>
<td>Total lifetime (21–83)</td>
<td>0.132</td>
<td>0.136</td>
<td>0.136</td>
<td>0.004 (3.0%)</td>
</tr>
</tbody>
</table>

Note: Old-age income is the sum of earnings from age 63–68 and pension income from age 63–83. Total lifetime income further includes earnings from age 21–62.

income than they would have had in the pre-reform pension regime. The degree of inequality increased, as reflected by the (small) shift in the Lorenz curves (panel c). The rise in inequality was driven by the redistribution of pension income. The difference between the Lorenz curves with and without labor supply responses taken into account is hardly discernable.

Table 3, panel A, summarizes the estimated distributional impacts in terms of Gini coefficients. Our primary interest lies in how the pension reform affected the old-age (age 63–83) income inequality, as reflected in the sum of labor earnings and pension income over the remaining lifetime from age 63. For comparison, we also compute Gini coefficients for prime-age (age 21–60) earnings (which were not affected by the reform), for labor earnings during the early retirement window (age 63–67), and for total lifetime income (age 21–83).

The Gini coefficient of 0.181 for prime-age earnings is relatively low. This reflects that the private sector workers qualifying for early retirement, both before and after the reform, generally had long and stable careers. This Gini coefficient compares to 0.264 for all workers employed at age 60, and 0.353 for all residents at age 60 regardless of their employment status (both based on the same birth cohorts).

By contrast, the Gini coefficient for earnings during the early retirement period was as high as 0.581 for the pre-reform cohorts in our sample, and it fell to 0.464 for the post-reform cohorts when it became more common to continue working. The corresponding numbers for the total population employed at the age of 60 are 0.510 and 0.471, for the pre- and post reform cohorts respectively (and 0.548 and 0.507 for all residents of the same cohorts).

The Gini coefficients for old-age income, which mainly consists of pension income (cf. Table 2), are even lower than those of prime-age income. This is because of a minimum pension level and a ceiling on the annual accrual of pension rights. The Gini coefficient, however, rose from 0.129 to 0.157 due to the redistribution of pension income, given the pre-reform labor supply patterns. Adding in the labor supply re-

---

9 As the pension system is based on a common life expectancy for everyone, we use the same assumption in our calculations, implying a life expectancy of 83 years.
sponsors does not change the Gini coefficient, so that the overall rise in the Gini coefficient is estimated to 0.028 (21.4%). To put this number into perspective, the transfer from the top to the bottom decile of the old-age income distribution in the post-reform sample required to reverse this increase in inequality would amount to approximately 10% of the overall old-age income in the top decile. However, the influence of this rise in old-age income inequality on the overall inequality in lifetime (21–83) income is moderate. We estimate that the Gini coefficient characterizing the distribution of total lifetime incomes rose by 0.001 (0.6%).

Panels B and C of Table 3 report corresponding inequality metrics separately for men and women. The reform-initiated rise in within-gender inequality was similar in magnitude as the rise in overall inequality. Yet, while the degree of inequality tends to be smaller among women, the rise in old-age income inequality caused by the reform was larger.

7. Conclusion

The usage of real (non-deferrable) retirement earnings tests (RET) in a pension system causes pension entitlements to be disproportionally allocated to people who retire early. As there is a strong social gradient in the preferred timing of retirement, such that workers with good jobs and high earnings typically wish to retire later than workers with bad jobs and low earnings, a retirement earnings test implies a more equal distribution of old-age income. However, when we take into account that the earnings test is likely to affect the labor supply of workers with different occupations and wage rates differently, it is no longer obvious how a RET ultimately affects the old-age income distribution. If low-wage workers respond sufficiently stronger to the work disincentives embedded in the RET than do high-wage workers, it is in principle possible that the reduction in labor earnings caused by RET outweighs the gain associated with higher pensions for those who retire in any case, such that a removal of RET actually reduces old-age income inequality.

Exploiting a comprehensive pension reform in Norway, we have examined the effects of RET on labor supply as well as on overall income inequality by comparing adjacent birth cohorts exposed to fundamentally different early retirement systems from age 62 to 67. We find that removal of the real RET which applied for a large segment of the Norwegian workforce until 2011, raised the labor supply over the whole 5-year early retirement period by approximately 6 h per week, or 42%. Although we identify considerable labor supply responses at all earnings levels, we find that the estimated effect sizes follow a hump-shaped pattern with respect to the prime-age (age 21–60) earnings distribution. The estimated labor supply effects of RET-removal vary from 5.4 h per week (42.6%) for the bottom decile, up to a maximum of 7.3 h (51%) for the 7th decile and then down again to 4.3 h (21.9%) for the top decile.

While the redistribution of pension wealth from early to late retirees implies that RET removal did increase inequality considerably, it turns out that the structure of the estimated labor supply responses had little effect on inequality. Adding up the direct effects (given the pre-reform distribution of employment and work hours) and the effects operating through changes in labor supply, we estimate that old-age income inequality, as measured by the Gini coefficient, rose by approximately 21% as a result of RET removal.

The findings reported in this paper suggest that policy makers face a particularly challenging tradeoff between efficiency and equity in the design of early retirement systems. The large labor supply responses that followed from the RET removal indicate considerable efficiency gains. Before the reform, Norwegian elderly workers could be subjected to real tax rates (incorporating the earnings test) between 80 and 100%. According to the findings in this paper, this instigated workers to leave the labor market in large numbers, despite the fact that many of them would have preferred to work with take-home wages somewhat closer to the true value of their labor. The RET essentially drives a huge wedge between the employer’s wage costs and the workers net pay, discouraging work even when its social value by far exceeds the private value of the forgone leisure. Thus, the RET appears to be a very expensive way of achieving a more equal income distribution.

To sum up: The removal of the retirement earnings test in the Norwegian early retirement system led to considerable increases in both labor supply (and economic efficiency) and in old-age income inequality. If the rise in income inequality is considered undesirable, a natural question to ask is whether it is possible to design the pension system such that it achieves the preferred redistribution of old-age incomes, but without incentivizing inefficient early retirement and thus imposing large welfare losses on the economy. Within the context of an actuarially fair early retirement system, this can be done by redistributing pension wealth toward workers with low prime-age earnings, i.e., by making the whole pension system more progressive, or by redistributing it toward occupational groups associated with early labor market exit on average. Alternatively, given that there is a positive correlation between life expectancy and the prime-age earnings level, it is possible to achieve a more egalitarian distribution of old-age income simply by distributing parts of the pension wealth in the form of time-limited (e.g. 10 or 15 years) rather than lifelong annuities, as is already the practice in some of Norway’s occupational pension schemes.

Appendix A. Eligibility in AFP-schemes

Eligibility for the old AFP schemes was determined in part by earnings-history and in part by employment at the time of take-out. The earnings-requirements consisted of three parts that all needed to be satisfied:

- Pensionable income above 1B in the take-out year and in the previous year.\(^\text{10}\)
- Pensionable income earnings above 1B in at least 10 years from age 50 (Last-10-rule).
- Average earnings above 2B in the 10 years with highest earnings after 1967 (Best-10-rule).

Pensionable income consists of wage earnings, self-employment earnings, and some temporary social insurance transfers (sick pay, unemployment insurance, and temporary disability insurance). In addition to the earnings-requirements, the individual has to be employed at the time of first take-out. Furthermore, one of these two conditions should be satisfied:

- Employment in the same private sector firm (with an AFP-scheme) in the last 3 years.
- Employment in a private sector firm (with an AFP-scheme) in the last 5 years.

\(^{10}\) B is the so-called Basic Amount (Grunnbeløpet) in the Norwegian pension system, currently (2019/2020) equal to approximately NOK 100,000 (\(\approx \$10,000\)), and annually adjusted in line with aggregate wage growth.
The eligibility criteria of the new private sector AFP are similar to those of the old scheme. As before, they consist of three parts, namely:

- An earnings-requirement (evaluated at the time of take-out)
- An employment requirement (evaluated at the time of take-out)
- An affiliation requirement (evaluated when turning 62)

The earnings-requirement is less strict than that of the old AFP scheme, since the Last-10 and the Best-10 rules no longer apply. Thus, the only requirement is that earnings at the time of take-out must exceed LB on an annual basis and that earnings in the preceding year must exceed the average of B in that year. The second requirement states that, in order to qualify for AFP, an individual must be “genuinely” employed in a company affiliated with the AFP-scheme at the time of take-out and must have been so in the previous 3 years. In order to qualify as “genuine” employment, the position should correspond to at least 20% of full-time, and it should represent the primary occupation and source of income. Finally, the affiliation requirement states that the individual must have been covered by the private sector AFP-scheme for at least 7 of the previous 9 years when turning 62. This replaces the requirement of affiliation in the previous 5 years applying in the old scheme. In order not to affect the cohorts close to retirement in 2011, this is implemented gradually. For the cohorts analyzed in the present paper, the requirement was 3 out of the last 5 years.

Appendix B. Estimation results for additional ages

See Figs. B1–B4
Appendix C. Reform effects by gender

See Fig. C1 and C2
Fig. C1. Estimated effects on average labor market outcomes age 63–67 by deciles based on alternative socioeconomic indicators. Men. Note: The point estimates (with 95% confidence intervals) indicate the effects on the 5-year average outcomes, measured over the calendar years in which individuals reach the ages of 63–67. The dotted lines indicate the average estimated effects for the total samples. See the note to Table 1 for a description of how we have defined and computed social class, life expectancy, and sick leave.
Fig. C2. Estimated effects on average labor market outcomes age 63–67 by deciles based on alternative socioeconomic indicators. Women. Note: The point estimates (with 95% confidence intervals) indicate the effects on the 5-year average outcomes, measured over the calendar years in which individuals reach the ages of 63–67. The dotted lines indicate the average estimated effects for the total samples. See the note to Table 1 for a description of how we have defined and computed social class, life expectancy, and sick leave.
References


Chapter 2:

Local labor demand and participation in social insurance programs
Local labor demand and participation in social insurance programs

Asbjørn Goul Andersen, Simen Markussen*, Knut Roed1

The Ragnar Frisch Centre for Economic Research, Gaustadalléen 21, Oslo 0349, Norway

A R T I C L E   I N F O
JEL codes:
J21
J58
J65
H55

Keywords:
Disability insurance
Program substitution
Shift-share analysis
Unemployment

A B S T R A C T
Based on administrative data from Norway, we explore the “gray area” between the roles of unemployment- and temporary disability-insurances by examining how participation in these two program types is affected by local labor demand conditions. Local labor demand is identified by means of a shift-share instrumental variables strategy, where initial local industry-composition is interacted with subsequent national industry-specific employment fluctuations. Our results indicate that local labor demand has a large negative effect on the propensity to claim disability insurance, which, for some groups, is remarkably similar to its effect on the propensity to claim unemployment insurance. Based on this finding, we question whether it is meaningful to maintain a sharp distinction between these two programs.

1. Introduction
Social insurance programs are typically designed such that they distinguish sharply between unemployment and disability as the foundation for claims. Is this distinction meaningful? For the majority of shorter spells of unemployment or sickness, the answer is probably yes. Most unemployment insurance claims reflect labor market frictions – it simply takes some time for persons who have become unemployed to find a good job match. And most sickness insurance claims arise due to some short-term ailment with no consequences for future employment opportunities. However, as social insurance spells become longer, the ultimate causes behind the claims often become more ambiguous. A person may be unemployed because, e.g., a musculoskeletal disease or a light mental disorder makes it difficult to compete for jobs. And a person may be considered disabled because expected productivity is too low to ensure realistic job opportunities. Long-term social insurance claims may also result from a combination of several labor market barriers, and although a claimant is declared either unemployed or disabled, (s)he may in reality be unemployed with respect to one job, disabled with respect to another, and perhaps unwilling with respect to a third. Health problems may of course make it difficult to perform some kind of tasks, while being irrelevant for others.

Existing empirical evidence indicates a significant degree of substitution between unemployment- and disability-related social insurance program utilization (Black et al., 2002; Autor and Duggan, 2003; Rege et al., 2009; Bratsberg et al., 2013; Maestas et al., 2015, 2018; Charles et al., 2018), and points to a considerable remaining work capacity among marginal disability insurance claimants (Maestas et al., 2013; Kostel and Mogstad, 2014; Borghans et al., 2014; French and Song, 2014). The probability of becoming a disability benefit claimant rises sharply in response to (exogenous) job loss. And although the positive relationship between layoff and disability risk to some extent reflects a genuine adverse health effect of job loss, the impacts identified in the empirical literature are simply too large to make this plausible as the sole explanation. In a recent US study, Maestas et al. (2018) estimate that 8.9% of all awards of Social Security Disability Insurance (SSDI) benefits during 2008–2012 was directly induced by the Great Recession. Based on Norwegian administrative data merged with records on mass layoffs identified from bankruptcy court proceedings, Bratsberg et al. (2013) estimate that men’s risk of claiming permanent disability benefits over the next few years more than doubles in response to a job loss. And conditional on having been laid off, the probability of becoming a disability benefit claimant rises steeply with the local rate of unemployment.

Some of the effects identified in the literature are likely to be context-dependent. For example, as a result of individual job loss, it is probable that health problems that were tolerated within an existing employment relationship become a barrier in a search for new employment. As pointed out by Autor and Duggan (2003), job displacement can be viewed as a negative shock to the value of continued labor market participation. Empirical evidence from Norway also confirms that displacement leads to significant earnings losses (Huttunen et al., 2011). Hence,

* Corresponding author.
E-mail address: simen.markussen@frisch.uio.no (S. Markussen).

1 This research has received support from the Norwegian Ministry of Labor and Social Affairs through the project “Unemployment in disguise.” Administrative registers made available by Statistics Norway have been essential. We thank the Editor and an anonymous referee for constructive comments and suggestions.

https://doi.org/10.1016/j.labeco.2019.101767
Received 4 July 2019; Received in revised form 12 September 2019; Accepted 16 September 2019
Available online 21 September 2019
0927-5371/© 2019 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license.
(http://creativecommons.org/licenses/by-nc-nd/4.0/)
while an existing job is preferred over inactivity, it is possible that a disability benefit application is preferred over search for new employment.

In the present paper, we explore the gray area between unemployment and disability in more detail by examining how the participation in different types of social insurance programs and subsequent labor market outcomes are causally affected by local employment opportunities. Rather than focusing specifically on persons exposed to individual shocks, such as a job loss, we study the influence of labor demand on program participation propensities for all adult individuals. In addition, we examine how the sensitivity of program participation to local labor demand fluctuations varies with respect to initial labor market state.

Norway is a country with relatively large fractions on disability-related social insurance programs, but relatively few on unemployment-related programs. Over the past decades, there has also been a systematic shift in the caseload away from unemployment-programs toward disability-programs. These points are clearly illustrated in Fig. 1, which shows the fractions of the adult population in Norway claiming the two types of benefits year-by-year since 1992. Particularly during the 1990s, there was a considerable increase in disability-related social insurance claims accompanied by a decline in unemployment-related claims. And based on the most recent numbers, there are now more than four persons on disability insurance for each person on unemployment insurance.

Empirical evidence indicates that whether a given labor market problem is interpreted by the social insurer as a health problem or as an unemployment problem may have real consequences in terms of later labor market outcomes, as unemployment programs tend to be less generous and also much more activation-oriented than disability programs. For example, Schreiner (2019) shows that a local social insurance office's overall tendency to interpret youth problems as health-related rather than unemployment-related has a considerable negative impact on the youths' future labor market outcomes.

In the present paper, we use Norwegian administrative register data to empirically assess the influence of local labor demand conditions on unemployment- and disability-related social insurance claims, respectively. To do this, we divide the country into commuting zones, and examine how the caseloads of the two program types are associated with local labor demand conditions based on variation across commuting-zone-by-year cells. To represent a source of exogenous variation in local labor demand, we use a Bartik-type shift-share instrument that interacts the initial local structure of employment across industries at various points in time, with the subsequent national fluctuations in industry-specific employment. In relation to the existing literature, we make two novel contributions. The first is that we identify the influence of labor demand fluctuations for representative populations, without relying on large individual or aggregate economic shocks; hence our results should score high on external validity. The second is that we offer a direct comparison of the influence that labor demand exerts on the caselogs of disability- and unemployment-related social insurances. This gives us a natural scale against which the effects on disability insurance program participation can be measured.

Our findings confirm that there is indeed a considerable gray area between unemployment—and disability-related insurance claims. Although the impact of local labor demand conditions on the probability of claiming unemployment-related economic support is larger than the corresponding impact on the probability of claiming disability-related support, the latter is far from negligible, particularly when we take into account that transitions into disability insurance tend to be highly persistent. For example, considering the population of newly employed workers, we estimate that the fraction claiming an unemployment-related benefit three years later decreases with 0.83 percentage points for every demand-initiated percentage point increase in the overall local employment rate, while the fraction claiming a disability-related benefit decreases by 0.25 percentage points. Conversely, having already entered unemployment or disability insurance, the same one-percentage point increase in local labor demand is estimated to increase the fraction having returned to work after three years by 3.0 percentage points for unemployment entrants and by 1.9 percentage points for disability insurance entrants. Hence, the influence of labor demand is considerable for the caseloads of both programs.

2. Institutions

The Norwegian social insurance system makes a distinction between unemployment-related and health/disability-related needs for income support; see Table 1. Unemployed individuals may claim unemployment insurance (UI) if past earnings exceed a certain threshold, or means-tested social assistance (SA); in both cases conditional on active job search and willingness to accept any suitable job offer. If deemed to be in need of additional qualification or placement services for reasons other than a health problem, it is also possible to participate in active labor market programs (ALMP) or in a more comprehensive “qualification program” (QP) offering a fulltime activity with some income support. Unemployment insurance provides a replacement rate of 62.4% up to an earnings level corresponding to approximately 108% of average

Fig. 1. Fraction of adult population with health-related and unemployment-related benefits by the end of each year 1992–2017. Source: Kann and Sutterud (2017, updated in 2018).
full-time-full-year earnings in Norway; see Table 1. UI benefits are conditioned on the unemployment spell being involuntary, however, and if a UI applicant quit a previous job voluntarily or was fired for cause, there is a 12-week embargo period on UI entitlements.

Persons who are in need of income support due to disability or other health problems may claim temporary or permanent disability insurance (DI). For employees, there is first a one-year entitlement to sick-pay (with 100% replacement), and during this period it is also illegal to fire the worker with reference to the sickness (employment protection regulations apply). After one year of absence, it is allowed to fire a worker who is unable to return to regular work due to sickness. It is then possible for the worker to apply for temporary or permanent disability benefits, with a typical replacement ratio around 66%. Persons who are not employed can apply for disability insurance directly, and there is no requirement of previous employment either. For persons without previous work experience, the benefit level is set to a fixed minimum level; see Table 1. For all applicants, the precondition is that the work capacity is reduced by at least 50% as a direct consequence of disability/impairment. This must be certified by an authorized physician, but the final decision is made by the social security administration (SSA). In most cases, DI claimants will first be enrolled into the temporary disability insurance program (TDI), which (currently) has a maximum duration of three years. During this period, various rehabilitation measures will be considered and possibly tried out. When TDI benefits are exhausted, many claimants move on to the permanent disability insurance (PDI) program, from which there is almost no prospects for returning fully to the labor market. For a more thorough description of the Norwegian DI system, see Fevang et al. (2017).

At the face of it, these insurances thus cover income losses caused by very different circumstances. However, with respect to the DI eligibility assessment of whether or not the work capacity is reduced by at least 50% due to health problems, the legislation allows the SSA to take the applicant’s current realistic work opportunities into account. This represents a possible channel whereby labor demand conditions may influence the assessment of disability insurance eligibility. Schreiner (2019) presents evidence that there is considerable room for caseworker judgement, and that screening practices vary considerable across time and space.

It is notable that while eligibility to unemployment insurance is conditional on (and proportional to) previous labor earnings, disability insurance can be claimed even without previous work experience. For disability claimants, there is also a minimum benefit level, currently corresponding to approximately 36% of average full-time-full-year earnings in Norway. Given the apparent scope for physician and caseworker judgment regarding the assessment of the reduced work capacity, it appears plausible that the assignment of individuals to the different programs to some extent is influenced by the degree of economic coverage they provide.

3. Data and descriptive evidence

Our empirical analysis is based on administrative registers covering all residents in Norway over the period from 1999 through 2016. The primary purpose of our analysis is to identify and estimate the causal influence of labor demand conditions on the probability of claiming unemployment-related and disability-related social insurance benefits. In order to do that, we need exogenous variation in labor demand conditions. Such variation clearly exists across local labor markets as well as over time. However, it is not generally observed. Natural candidates for representing labor demand fluctuations in an empirical model are the local employment or unemployment rates (or other measures of labor market tightness). However, these are determined through the intersection of demand and supply; hence, they cannot be used directly as explanatory variables in a model intended to isolate the influence of labor demand. Across space and time, there will be a sort of mechanic relationship between the rates of social insurance program participation

### Table 1

<table>
<thead>
<tr>
<th>Income support programs targeted at unemployed job seekers and persons with disability or health problems in Norway.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Income support programs</strong></td>
</tr>
<tr>
<td>Unemployment-related insurance programs (UI)</td>
</tr>
<tr>
<td>Social assistance (SAS)</td>
</tr>
<tr>
<td>Qualification program (QP)</td>
</tr>
<tr>
<td>Active labor market program (ALP)</td>
</tr>
<tr>
<td>Health/disability-related insurance programs</td>
</tr>
<tr>
<td>Temporary disability insurance (TDI)</td>
</tr>
<tr>
<td>Permanent disability insurance (PDI)</td>
</tr>
<tr>
<td>Note: B is short for the “Basic amount”, which is an important monetary parameter in the Norwegian social insurance system. In 2019 it is approximately NOR 100,000, which is around 18% of average full-time-full-year earnings in Norway; see Table 1. The table describes the current (2019) system. Some of the parameters have been subject to changes during the period covered by this paper.</td>
</tr>
</tbody>
</table>

---
and the rate of employment (as these states to some extent are mutually exclusive); but without additional information we cannot identify the direction of causality. For example, if we observe that a local labor market at some point in time has a particularly low employment rate and a high rate of disability program participation, we still don’t know whether it is the low employment rate that causes the high disability rate or vice versa.

To deal with this simultaneity problem, we will use a shift-share strategy in which we interact national industry-specific employment fluctuations with some initial spatial differences in industry composition. Instruments of this type have been used frequently in the literature; see, e.g., Bartik (1991), Blanchard and Katz (1992), Bound and Holzer (2000), Autor and Duggan (2003), and Bartik (2015). They isolate the employment fluctuations following directly from the national expansion and contraction of particular industries, caused by, say, changes in technology, trade liberalization (or exchange rate fluctuations), public expenditures, or consumer demand. The central identifying assumption is then that the initial industry shares do not predict future outcomes through other channels than those reflected in the national fluctuations; see Goldsmith-Pinkham et al. (2019). To operationalize this empirical strategy in the context of our data, we need to structure the data in terms of base-years (used to define the initial industry composition) and outcome-years. As explained in more detail below, we will in the main part of the analysis do this by allowing a three-year time-period between base-years and outcome-years. In addition, we need to define local labor markets. Based on Bhuller (2009), we divide the country into 46 such local markets, or commuting zones. The overall variation in labor demand exploited in our analysis thus comes from 690 combinations of 15 different base-years and 46 different commuting zones. We also need to assign all employees in Norway to particular industries. Such information is available in administrative registers, based on a five-digit industry code. In total, there are 648 different industries in Norway. However, many of these are very small, and perhaps located only in a few commuting zones, hence, we aggregate industries such that all industry-categories have at least 5,000 employees on average at an annual basis. This is done by first including all five-digit codes with at least 5,000 employees, then doing the same for four-digit codes, and so on. As a result, we end up with 171 unique industries.

Fig. A1 in Appendix A provides a compact description of the longitudinal and cross-sectional variation in employment across industries in Norway. With a prominent exception for the oil and gas industry, there has been considerable decline in the employment share of production industries over the 1999–2016 period. As the importance of these industries also vary a lot across commuting zones, this is an important exogenous source of variation in labor demand conditions. The largest growth in employment has come in the construction industries and in the service sectors related to health and education. For these industries, the variation in employment shares is much smaller, yet far from negligible.

To study the influence of labor demand on social insurance program participation, we combine information from several administrative registers to assign unique monthly labor market states to all adult (age 18–61) residents in Norway. Our analysis focuses on four states; i.e., employment, participation in a disability-related social insurance program, participation in an unemployment-related program, and education, respectively. In order to characterize each person’s main economic activity, the states are defined as mutually exclusive. In cases where people apparently have belonged to multiple states within the same month, uniqueness is achieved applying a hierarchy, which ranks states after their presumed distance to the labor market. This implies that health-related benefit claims are prioritized over unemployment, which is again prioritized over education, and finally employment. This hierarchy has the additional advantage of prioritizing data sources where the monthly information is considered most reliable. The definition of labor market states is described in more detail in Appendix B.

We construct two types of datasets. The first contains the complete stock of individuals, with a fine-grained record of labor market states by January each year. These data are again divided into different groups depending on initial state. The second type of dataset contains, for each year, new entrants into either employment, unemployment-related income support, or temporary disability insurance. A new entrant to a particular state is defined as being in that state in a given month, while not having belonged to that state during the previous three months. By focusing on entrants, we direct attention directly to those whose subsequent labor market performance presumably is most sensitive with respect to labor demand conditions.

The structure of the datasets is illustrated in Table 2. In total, there are almost 38 million observations divided between 3.6 million individuals, and 66% of these observations start out with employment in the base-year; see Column I. Among them, 87% are still in employment three years later, 2% have become unemployed, and 3% have become disabled; see Column IV. Employment is less stable for the newly employed (Column V), and their risk of becoming unemployed or disabled is also much higher.

Although the data contain information about individual outcomes, the variation in labor demand conditions comes from the 690 different combinations of base-years and commuting zones. Hence, most of the analysis can be done based on aggregates computed for each commuting-zone-by-year cell. Before we present our empirical model, we provide a descriptive picture of the relationship between fractions belonging to the three key states of employment, unemployment, and disability insurance in the respective base-years. For this purpose, we build on the data containing all adults described in Table 2, Column I. The upper two panels of Fig. 2 first show that there is a strong negative relationship between local employment rates and both unemployment (panel (a)) and disability-related (panel (b)) insurance claims. This is not surprising, given that such claims by construction implies non-employment. The relationships are not entirely mechanistic, though, as approximately 20% of the population does not belong to any of these states of unemployment, disability, and employment; see Table 2. The four lower panels illustrate that the cross-sectional variation (panels (c) and (d)) in all three rates are much larger than the longitudinal variation (panels (e) and (f)). They also indicate that while longitudinal variation in employment is most strongly (negatively) associated with participation in unemployment-related insurance programs, its cross-sectional variation is most strongly associated with participation in disability-related programs.

Panel (a) in Fig. 3 then focuses more directly on the relationship between the rates of unemployment-related and disability-related claims. At the face of it, there is a positive relationship between these two rates at the commuting-zone-by-year level; see the upwards-sloping stapled regression line. However, when we instead look at the relationships between the two caseloads among local areas with similar employment rates, a completely different pattern emerges. Then, there is a conspicuous negative relationship between the rates of unemployment and disability. Again, it is worth noting that this pattern is not purely mechanistic. To illustrate this point, panels (b) and (c) in Fig. 3 show the corresponding relationships between the respective local fractions of social insurance program participation and the fraction belonging to “other” non-employment states, based on exactly the same grouping of employment rates as used in panel (a). In these graphs, the systematic and tidy patterns displayed in panel (a) appear to be completely absent. Although this descriptive evidence is far from conclusive, it may point toward two suppositions; first, that unemployment and disability program participation are driven by some common determinant (e.g., cyclical fluctuations in the level of labor demand), and second, that, given the level of labor

2 “Other” non-employment states include education, homemaking, periods spent outside the country, and inactivity.
demand, there is an important element of substitution between the two program types.

4. Empirical strategy

To establish more conclusive evidence regarding the “gray area” between unemployment and disability insurances, we now set up a more formal statistical model aimed at identifying and estimating the influence that labor demand actually has on the two caseloads. Using the seven different samples described in Table 2, we examine four different outcomes, all defined at the commuting-zone-by-year level: i) the fraction in employment, ii) the fraction in unemployment-related insurance, iii) the fraction in disability-related insurance, and iv) the fraction in education.

To describe our regression models, we need some notation. Let the subscript $b$ indicate the base-year and let $t$ indicate the outcome-year. In the stock sample, the base-year observations are defined in terms of the January records each year (1999–2013), whereas in the entrant samples it is defined in terms of records corresponding to the month of entry. The outcome-year observations are in the main part of our analysis measured exactly three years later (2002–2016). However, in Appendix D, we also present results for outcome-years measured from just one and up to seven years after the base-year. The subscript $z$ indicates commuting zone, which always refers to the commuting zone occupied in the base-year.

Let $y_{zt}^b$ be the fraction of the respective base-year population in commuting zone $z$ that belongs to a state $s$ in the outcome year $t$. Abstracting from the obvious problem that local labor demand is intrinsically unobserved, we would have liked to regress each outcome on the level of labor demand, while controlling for initial conditions and the composition of individuals under study; i.e.:

$$y_{zt}^b = Y_{zb} + X_{zb} + L_{D_{zt}} + \mu_{zt},$$

where $L_{D_{zt}}$ is a measure of local labor demand in commuting zone $z$ in the outcome-year $t$, $Y_{zb}$ is a vector containing the state-specific population shares measured in the base-year, including the baseline value of the dependent variable ($y_{zt}^b$), and $X_{zb}$ is a vector of average individual characteristics within the commuting zone’s base-year sample (gender, age, education, immigrant status, and earnings; all measured in (or prior to) the base-year).

Our primary interest lies in the impact of local labor demand; i.e., $L_{D_{zt}}$. However, as pointed out above, labor demand is unobserved. A natural proxy for labor demand is the overall employment rate in the commuting zone at the time of outcome measurement. However, in order to isolate the exogenous fluctuations due to variations in labor demand, we need a valid instrument; i.e., we need a variable that affects the local employment rate through a channel of labor demand, but otherwise satisfies an exclusion restriction with respect to Eq. (1). We use a Bartik instrument of the following kind

$$z_{zt} = \frac{\sum_{j=1}^n w_{ij}(L_{ij} - \bar{L}_i)}{N_{zb}},$$

where $w_{ij}$ is commuting zone $z$’s fraction of employees within industry $j$ in base-year $b$, $(L_{ij} - \bar{L}_i)$ is the total change in the number of employees in industry $j$ from the base-year to the outcome-year in the whole

---

**Table 2**

<table>
<thead>
<tr>
<th></th>
<th>The complete stock by January each year</th>
<th>Entrants to...</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All</td>
<td>Employed</td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>II</td>
</tr>
<tr>
<td>N</td>
<td>37 939 858</td>
<td>25 554 728</td>
</tr>
<tr>
<td>Female</td>
<td>0.49</td>
<td>0.44</td>
</tr>
<tr>
<td>Age</td>
<td>38.2</td>
<td>40.6</td>
</tr>
<tr>
<td>Educational attainment Compulsory</td>
<td>0.36</td>
<td>0.30</td>
</tr>
<tr>
<td>High school</td>
<td>0.33</td>
<td>0.35</td>
</tr>
<tr>
<td>College</td>
<td>0.30</td>
<td>0.33</td>
</tr>
<tr>
<td>Labor earnings last year (B)</td>
<td>4.4</td>
<td>6.0</td>
</tr>
<tr>
<td>Immigrant low-income country</td>
<td>0.10</td>
<td>0.09</td>
</tr>
<tr>
<td>Immigrant high-income country</td>
<td>0.01</td>
<td>0.00</td>
</tr>
<tr>
<td>State in base-year Employed</td>
<td>0.66</td>
<td>1.00</td>
</tr>
<tr>
<td>Unemployed</td>
<td>0.04</td>
<td>0.00</td>
</tr>
<tr>
<td>Disability insurance</td>
<td>0.09</td>
<td>0.00</td>
</tr>
<tr>
<td>In education</td>
<td>0.15</td>
<td>0.00</td>
</tr>
<tr>
<td>State in outcome-year (three years later) Employed</td>
<td>0.69</td>
<td>0.87</td>
</tr>
<tr>
<td>Unemployed</td>
<td>0.03</td>
<td>0.02</td>
</tr>
<tr>
<td>Disability insurance</td>
<td>0.12</td>
<td>0.03</td>
</tr>
<tr>
<td>In education</td>
<td>0.10</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Note: The entry of observation is person-year. All residents in Norway are included for each base-year 1999–2013, provided that they are between 18 and 58 years of age in the respective base-year (column I) and that they satisfy the initial state criteria indicated in the column heads (columns II–VII). In columns I–IV (the stock samples), base-year state is recorded in January each year. In columns V–VII (the entrant samples), base-year state is recorded in the month of entry. The state in the outcome-year is in both cases recorded exactly three years later. The unemployed and disability insurance states correspond to the categorization used in Table 1, such that unemployment comprises UI, SA, QP, and ALMP, and disability insurance comprises TDI and PDI.

---

3 To control appropriately for variations in initial conditions across commuting-zone-by-year cells, the vector $x_{zb}$ contains a more fine-grained state space than the outcome variables; i.e., the fractions belonging to i) full-time employment, ii) part-time employment, iii) self-employment, iv) parental leave, v) sick-pay, vi) unemployment insurance or participation in active labor market program, vii) social assistance or qualification program, viii) temporary disability insurance, and ix) permanent disability insurance (see the Appendix B for a more detailed description of the state-space). The vector $x_{zb}$ contains the following variables: i) the fraction of females, ii) the fraction with high-school (upper secondary) education, iii) the fraction with college/university education, iv) the fractions belonging to different 5-year age intervals, v) the fraction of immigrants from low-income countries, vi) the fraction of immigrants from high-income countries, and vii) average labor earnings in the year prior to the base-year.
country, and \( N_{z_t} \) is the size of the adult population in commuting zone \( z \) in the base-year.

The instrument \( z_{it} \) is thus the predicted change in the local employment rate from the base-year to the outcome-year, based on the national changes in the industry-specific employment patterns only, and measured relative to the size of the base-year population. Taken at face value, the instrument in (2) also incorporates national changes in the overall employment rate, which may stem from fluctuations in labor supply as well as demand. Hence, to ensure that the identifying information provided by the instrument encompasses the idiosyncratic changes related to industry-composition only, we will control for outcome-year fixed effects. In addition, we control for commuting-zone fixed effects to ensure that any stable correlation between the initial industry structure and labor supply behavior across commuting zones is not picked up by the instrument. Finally, since the local employment rate instrumented by \( z_{it} \) may deviate from the employment rate observed within the samples under study, we also control for the base-year value of the instrumented employment rate.

The baseline two-stage least squares (2SLS) models we estimate thus have the following form:

\[
y_{it} = \alpha_i + \delta_{it} + \beta y_{it} + \gamma x_{it} + \mu z_{it} + \xi_{it} + \epsilon_{it}. \tag{3}
\]

where \( \alpha_i \) are the year-fixed effects, \( \delta_{it} \) are the commuting-zone-fixed effects and \( \epsilon_{it} \) is the employment rate (age 25–60) in commuting zone \( z \) in
the base-year, and $\hat{e}_{yt}$ is the corresponding predicted employment rate for the outcome-year based on the first stage equation

$$e_{yt} = \hat{\phi}_r + \varphi_{yt} + \varphi_{yt} \times \beta_{yt} + \eta_{yt} + \mu + \zeta_{yt}.$$  

We estimate the model using the 690 commuting-zone-by-year observations, with weights reflecting the number of individual observations behind each data point.

While we build on this model in the presentation of results in the next section, we show in Section 6 and in Appendix E that the results are robust with respect to a number of alternative specifications. These include the use of alternative control variables (e.g., allowing for local linear time trends) and the use of a modified instrument where the influence of own commuting zone in national trends is removed (i.e., a “leave-out” Bartik instrument). They also include the use of individual data (instead of commuting-zone-by-year cells), which allows for the inclusion of individual-fixed effects. The robustness analysis also incorporates estimation of a different model, where the years used to compute initial industry weights are kept constant across different base-years, facilitating the inclusion of commuting-zone-by-weight-constuction-year fixed effects. Finally, we present a “placebo” analysis where we use past instead of future outcomes as dependent variables in the baseline model.

5. Main results

A critical precondition for this empirical strategy to work is that there is a sufficiently strong first stage; i.e., that the national fluctuations in industry-composition really have a substantial impact on local employment patterns. Fig. 4, panel (a) first assesses this graphically, by plotting the realized change in local employment rate $e_{yt}$ against its prediction...
The relationship indeed appears strongly positive. However, as argued above, in the model we need to control for both year-fixed and commuting-zone fixed effects to ensure that we isolate the influences of labor demand; hence the variation actually exploited in the model is the variation remaining after having controlled for these factors. This is illustrated in Fig. 4, panel (b). The relationship then becomes considerably weaker, but still positive.

Table 3 presents the first stage estimation results from Eq. (4). They show that the instrument is sufficiently strong for valid statistical inference within all the samples described in Table 2. Having confirmed sufficient strength of the instrument, we now turn to the main results; see Table 4. For comparison, we present corresponding ordinary least squares (OLS) estimates in Appendix C. In most cases, the 2SLS estimates are a bit larger than the OLS estimates. There are two reasons why OLS and 2SLS estimates may differ. The first is directly related to the simultaneity problem discussed above, i.e. that the residual in Eq. (3) is correlated with the local employment rate. As a particularly high employment rate may indicate some favorable labor supply developments in the region, this is likely to exaggerate the influence of labor demand. The second reason is that the observed employment rate is an imperfect measure of labor demand, and thus subjected to measurement error. This will tend to bias the OLS estimates toward zero. In our case, it appears that the latter source of bias in most cases dominates the former.

Returning to the 2SLS estimates in Table 4, Column I first provides the results obtained for the full stock sample. As expected, the estimated effect on the employment rate in the full sample is approximately equal to 1. This particular result is almost tautological, as the population behind this estimate is almost the same as the population behind the first stage. However, the estimates regarding the states that the higher employment rate substitutes for are of more substantive interest. We note that a 1-percentage point demand-driven increase in the local employment rate reduces the local unemployment rate by 0.68 percentage points, and the rate of disability insurance program participation by 0.23 percentage points. The estimates also indicate a slight reduction in the probability of being in education, but this effect is not statistically significant.

It may also be of some interest to see how the effects reported in Column I vary across different demographic and educational groups. To shed light on this, we have estimated the model separately for 4 different age groups and for 12 different combinations of age and educational attainment. For ease of comparison, we present the results from this exercise graphically; see Fig. 5. It is clear that the effects of labor demand fluctuations are largest for the young, and among them, there is a ten-
Table 4
Second stage estimates: Effects of local labor demand on the fractions belonging to different states in outcome-year. By initial state.

<table>
<thead>
<tr>
<th>Employment</th>
<th>Employment-related insurance</th>
<th>Disability insurance</th>
<th>Education</th>
<th># Observations</th>
<th># Person-years</th>
</tr>
</thead>
<tbody>
<tr>
<td>All I</td>
<td>Employed II</td>
<td>Unemployed III</td>
<td>Disabled IV</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.057***</td>
<td>0.712***</td>
<td>3.519***</td>
<td>0.614**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.115)</td>
<td>(0.144)</td>
<td>(0.363)</td>
<td>(0.260)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unemployment</td>
<td>–0.603***</td>
<td>–2.03***</td>
<td>0.056</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.096)</td>
<td>(0.544)</td>
<td>(0.098)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dis 3.519***</td>
<td>–0.231***</td>
<td>–1.028***</td>
<td>–1.028***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.065)</td>
<td>(0.373)</td>
<td>(0.345)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education</td>
<td>0.147</td>
<td>0.038</td>
<td>0.038</td>
<td>690</td>
<td>37 939 858</td>
</tr>
<tr>
<td></td>
<td>(0.115)</td>
<td>(0.096)</td>
<td>(0.050)</td>
<td>690</td>
<td>25 554 728</td>
</tr>
<tr>
<td></td>
<td></td>
<td>690</td>
<td>690</td>
<td>1 280 145</td>
<td>1 280 145</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3 791 759</td>
<td>3 791 759</td>
<td>4 469 054</td>
<td>4 469 054</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 919 349</td>
<td>1 919 349</td>
<td>688 913</td>
<td>688 913</td>
</tr>
</tbody>
</table>

Note: Each coefficient in this table is a result of a separate weighted 2SLS regression based on Eqs. (3) and (4). Standard errors are clustered at commuting zone. ***/*** indicates statistical significance at the 10/5/1 percent levels.

Fig. 5. Second stage estimates: Effects of local labor demand on the probability of belonging to different states in outcome-year. Complete stock sample by age and educational attainment.

Note: The age grouping is shown on the horizontal axis. The three educational groups indicated in the legend are defined as follows: Comp.edu: compulsory education only or incomplete high-school education; HS: high school (upper secondary) education; College: College/University degree. Each coefficient is a result of a separate weighted 2SLS regression based on Eqs. (3) and (4). Point estimates are shown with 95% confidence intervals.

dency that the effects are largest for those with least education. It is also for the uneducated young people that we see the strongest evidence that labor demand conditions influence disability program participation. Effects on continued education are almost exclusively concentrated among the young.

As it turns out, it is not primarily age or educational attainment per se that determines the size of labor demand effects, but rather the initial state, which is highly correlated with age and education. Moving on to the results that are conditioned on the initial state (Tables 3 and 4, columns II-IV), we note that the effects of labor demand are systematically larger for unemployed job seekers. For them, a 1 percentage point demand-driven increase in the local employment rate over the next three years is estimated to increase employment propensity by 3.5 percentage points (Column III). Most of this effect comes from reduced unemployment propensity (–2.2 percentage points). However, it is notable that the probability of having moved on to disability insurance is also reduced almost in line with the increase in labor demand (–0.86 percentage points). The much smaller effects estimated for those who al-

57
ready belonged to a disability state in the base-year (Column IV) reflect that the majority of them actually belonged to the state of permanent disability insurance, which tends to be an absorbing state in Norway. Yet, it is notable that the probability of remaining in a disability insurance state after three years fluctuates approximately one for one with demand-driven variations in the employment rate.²

Columns V-VII presents the estimation results for the three entrant (flow) samples; i.e., the group of people that had just become either employed, unemployed, or a temporary disability claimant in the base-year. Within all these entrant groups, the probability of employment three years later is highly dependent on local labor demand conditions. It is notable that the labor demand sensitivity of new entrants to unemployment- and disability-related programs is much more similar than it is for the two stocks. While a 1 percentage point demand-driven increase in local labor demand is estimated to raise the probability of being employed three years later by 3 percentage points for the newly unemployed, it raises it by 2 percentage points for the newly disabled.

The choice of a three-year distance between the base-year and the outcome-year is a bit arbitrary. It represents a compromise between ensuring appropriate local industry weights (which requires a relatively short distance) and ensuring sufficient variation in labor demand conditions (which requires a relatively long distance). For the complete stock sample, the choice of distance between base-year and outcome-year should not have any impact on point estimates, as the base-year is non-informative with respect to the initial state. For the samples that are conditioned on a particular state, on the other hand, the choice of distance is potentially substantively important, as longer distance attenuates the influence of the initial state. In Appendix D, we present complete estimation results for alternative choices of the outcome year, from one to seven years after the base-year. As expected, the estimates are quite stable for the complete stock sample, as well as for the samples that are based on initial states that on average tend to be persistent (the stock samples of employed and disability insurance claimants). For the other samples, there is a tendency for the estimated labor demand effects to be largest the closer in time the outcome is measured relative to the (precarious) initial state. This is particularly evident for the two unemployment samples, whose members are known to be looking for jobs in the base-year.

6. Robustness

To examine the robustness of our estimation results, we present, in Appendix E, complete results for outcomes measured three years after the base-year, based on five alternative specifications of the 2SLS model in Eqs. (3) and (4). First, we examine the sensitivity with respect to the inclusion of the control variables contained in \( \mathbf{x}_u \) and \( \mathbf{y}_u \) (mean individual covariates and the distribution of initial states), by estimating the model without any of these controls. This is of particular interest in relation to the models that are conditioned on an initial state, as the composition of entrants to the various labor market states may depend on labor demand conditions. By excluding/including individual controls, we can assess the results’ sensitivity with respect to this potential source of disturbance.

Second, we examine robustness with respect to the inclusion of regional trends in employment that are not driven by demand, but potentially correlated to initial industry weights. We do this by extending the baseline model to include commuting-zone-specific linear time trends. Third, as we in the baseline model have included each commuting-zone’s own employment in the national trends used to construct the

---

² When we estimate models separately for different age- and education groups conditional on initial state, the systematic relationship between the effects of labor demand and age/education illustrated in Fig. 5 disappears (not shown).
Bartik instrument, it could be argued that the national trends are not completely exogenous. It is possible to deal with this problem by using a “leave-out” Bartik instrument; i.e., an instrument where the national industry-specific employment trends are computed without including the focal commuting zone. However, the expansion of employment in one region may be causally related to contraction in another, e.g., because a large production unit has changed location. It is therefore not obvious which strategy provides the best foundation for causal analysis. We thus include a model built on a “leave-out” instrument in the robustness analysis. This leave-out instrument is constructed by substituting Eq. (2) with the following: 

$$z^*_{z} = \frac{\sum_{j=1}^{J} w_{zj} (I_{zj,2} - I_{zj,1})}{N_{z}},$$

where the $-z$ subscript indicates that the variable does not include commuting zone $z$.

Fourth, as we have estimated the model based on aggregate data (commuting-zone-by-year cells), it could be argued that we have not exploited individual data efficiently. In a robustness exercise, we thus use individual observations, allowing for a more flexible use of individual controls and initial states. The outcome variables then take the form of 0–1 (dummy) variables indicating whether or not the person belonged to the state in question in the outcome year, and standard errors are computed with a two-way cluster (individuals and region). The vector $x_{bh}$ is replaced by $\bar{x}_{bh}$, which contains the individual covariates, and the $y_{nh}$ is replaced by $\bar{y}_{nh}$, which contains dummy variables indicating the initial state for each person.

Fifth, based on individual data, we estimate a model with person-fixed effects. While this is relatively straightforward in the stock-samples, where most persons are included with 15 observations (one for each base-year), it is a bit more challenging in the entry samples, as many individuals do not experience more than one entry into a particular state. This implies that models with individual-fixed effects are estimated with considerable uncertainty for these samples.

As can be seen from Fig. E1 in Appendix E, the main message coming out of these exercises is that the results are indeed robust with respect to model specification. Although some of the point estimates vary slightly from model to model, none of the main results discussed above would have been substantively changed had we relied on a different version of the model.

A potential concern related to all the models based on Eqs. (3) and (4) is that the time variation in local industry weights within commuting zones may induce a simultaneity problem into the model, as these weights may be correlated to the error term in Eq. (3); confer the discussion in Goldsmith-Pinkham et al. (2019). It is obviously not possible to include commuting-zone-by-base-year dummy variables, as this would exhaust all the identifying information in the data. The stability of the results with respect to the inclusion of local time linear trends is reassuring in this respect. However, it is also possible to deal with this concern more directly; i.e., by keeping local industry weights constant across different base-years, and then include dummy variables for each combination of commuting zone and year of weight construction. The results from such a model are reported in Appendix Table E1, and they confirm robustness of our main findings also with respect to this specification.

As a final check on empirical strategy, we report in Appendix E the results from a placebo version of our baseline model, where we have substituted outcomes observed three years before the base-year for the outcomes observed three years after. By construction, labor demand developments in the three-year period after the base year cannot have had

---

Fig. E1. Robustness analysis. Estimated second stage coefficients with 95% confidence intervals. Note: The standard errors used to compute confidence intervals are clustered on region (in models with aggregate data) and on region and individuals (in the models with individual data).
Table C1
Ordinary Least Squares (OLS) estimates: Effects of local labor demand on the fractions belonging to different states in outcome-year. By initial state.

<table>
<thead>
<tr>
<th></th>
<th>All</th>
<th>Employed</th>
<th>Unemployed</th>
<th>Disabled</th>
<th>Employment</th>
<th>Unemployment</th>
<th>Temp. disability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I</td>
<td>II</td>
<td>III</td>
<td>IV</td>
<td>V</td>
<td>VI</td>
<td>VII</td>
</tr>
<tr>
<td>Employment</td>
<td>0.801***</td>
<td>0.687***</td>
<td>2.116***</td>
<td>0.565**</td>
<td>1.199***</td>
<td>2.214***</td>
<td>1.581***</td>
</tr>
<tr>
<td></td>
<td>(0.045)</td>
<td>(0.057)</td>
<td>(0.204)</td>
<td>(0.083)</td>
<td>(0.111)</td>
<td>(0.217)</td>
<td>(0.201)</td>
</tr>
<tr>
<td>Unemployment-related</td>
<td>–0.471***</td>
<td>–0.433***</td>
<td>–1.223***</td>
<td>0.011</td>
<td>–0.633**</td>
<td>–1.409***</td>
<td>–0.001</td>
</tr>
<tr>
<td>insurance</td>
<td>(0.073)</td>
<td>(0.072)</td>
<td>(0.253)</td>
<td>(0.041)</td>
<td>(0.080)</td>
<td>(0.261)</td>
<td>(0.098)</td>
</tr>
<tr>
<td>Disability insurance</td>
<td>–0.277***</td>
<td>–0.177</td>
<td>–0.693**</td>
<td>–0.769***</td>
<td>–0.162***</td>
<td>–0.680***</td>
<td>–1.603***</td>
</tr>
<tr>
<td></td>
<td>(0.057)</td>
<td>(0.042)</td>
<td>(0.029)</td>
<td>(0.150)</td>
<td>(0.033)</td>
<td>(0.092)</td>
<td>(0.353)</td>
</tr>
<tr>
<td>Education</td>
<td>0.002</td>
<td>–0.036</td>
<td>0.040</td>
<td>0.049*</td>
<td>–0.244**</td>
<td>0.039</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>(0.042)</td>
<td>(0.041)</td>
<td>(0.063)</td>
<td>(0.020)</td>
<td>(0.106)</td>
<td>(0.066)</td>
<td>(0.061)</td>
</tr>
<tr>
<td># Observations</td>
<td>690</td>
<td>690</td>
<td>690</td>
<td>690</td>
<td>690</td>
<td>690</td>
<td>690</td>
</tr>
<tr>
<td># Person-years</td>
<td>37 939 858</td>
<td>25 554 728</td>
<td>1 280 145</td>
<td>3 791 759</td>
<td>4 469 054</td>
<td>1 919 349</td>
<td>688 913</td>
</tr>
</tbody>
</table>

Note: Each coefficient in this table is a result of a separate weighted OLS regression based on Equations (3), with the actual employment rate (age 25–60) in the commuting zone in the outcome year substituted for the predicted rate. Standard errors are clustered at commuting zone. ‘*/’/*** indicates statistical significance at the 10/5/1 percent levels.

Table E1
Estimated effects of local labor demand on the probability of belonging to different states in outcome-year. By initial state. Alternative model.

<table>
<thead>
<tr>
<th></th>
<th>All</th>
<th>Employed</th>
<th>Un-employed</th>
<th>Disabled</th>
<th>Employment</th>
<th>Unemployed</th>
<th>Temp. disability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I</td>
<td>II</td>
<td>III</td>
<td>IV</td>
<td>V</td>
<td>VI</td>
<td>VII</td>
</tr>
<tr>
<td>Employment</td>
<td>0.856***</td>
<td>0.788***</td>
<td>3.937***</td>
<td>0.821**</td>
<td>1.355***</td>
<td>3.345***</td>
<td>3.103***</td>
</tr>
<tr>
<td></td>
<td>(0.090)</td>
<td>(0.098)</td>
<td>(0.924)</td>
<td>(0.210)</td>
<td>(0.475)</td>
<td>(0.640)</td>
<td>(0.969)</td>
</tr>
<tr>
<td>Unemployment-related</td>
<td>–0.681***</td>
<td>–0.736***</td>
<td>–1.551***</td>
<td>0.277**</td>
<td>–0.758**</td>
<td>–2.127**</td>
<td>–0.431**</td>
</tr>
<tr>
<td>insurance</td>
<td>(0.082)</td>
<td>(0.075)</td>
<td>(0.622)</td>
<td>(0.128)</td>
<td>(0.217)</td>
<td>(0.431)</td>
<td>(0.224)</td>
</tr>
<tr>
<td>Disability insurance</td>
<td>–0.352***</td>
<td>–0.047***</td>
<td>–1.030***</td>
<td>–1.523***</td>
<td>–0.267**</td>
<td>–0.790**</td>
<td>–1.366**</td>
</tr>
<tr>
<td></td>
<td>(0.095)</td>
<td>(0.051)</td>
<td>(0.521)</td>
<td>(0.493)</td>
<td>(0.101)</td>
<td>(0.243)</td>
<td>(1.368)</td>
</tr>
<tr>
<td>Education</td>
<td>–0.162(1.02)</td>
<td>0.068(0.079)</td>
<td>–0.432(0.270)</td>
<td>–0.029(0.082)</td>
<td>–0.082(0.309)</td>
<td>0.122(0.233)</td>
<td>–0.022(0.236)</td>
</tr>
<tr>
<td></td>
<td>(2990)</td>
<td>(2990)</td>
<td>(2990)</td>
<td>(2990)</td>
<td>(2990)</td>
<td>(2990)</td>
<td>(2990)</td>
</tr>
<tr>
<td># Observations</td>
<td>690</td>
<td>690</td>
<td>690</td>
<td>690</td>
<td>690</td>
<td>690</td>
<td>690</td>
</tr>
<tr>
<td># Person-years</td>
<td>37 939 858</td>
<td>25 554 728</td>
<td>1 280 145</td>
<td>3 791 759</td>
<td>4 469 054</td>
<td>1 919 349</td>
<td>688 913</td>
</tr>
</tbody>
</table>

Note: Each coefficient in this table is a result of a separate 2SLS regression based on Eqs. (3) and (4). Standard errors are clustered on commuting-zone. ‘*/’/*** indicates statistical significance at the 10/5/1 percent levels.

Table E2
Placebo analysis: Estimated effects of local labor demand on the probability of belonging to different states three years before the base-year. By state in base-year.

<table>
<thead>
<tr>
<th></th>
<th>All</th>
<th>Employed</th>
<th>Un-employed</th>
<th>Disabled</th>
<th>Employment</th>
<th>Unemployed</th>
<th>Temp. disability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I</td>
<td>II</td>
<td>III</td>
<td>IV</td>
<td>V</td>
<td>VI</td>
<td>VII</td>
</tr>
<tr>
<td>Employment</td>
<td>0.069(0.244)</td>
<td>–0.172(0.183)</td>
<td>–0.204(0.602)</td>
<td>–0.069(0.407)</td>
<td>–0.412(0.328)</td>
<td>0.264(0.620)</td>
<td>0.496(0.980)</td>
</tr>
<tr>
<td></td>
<td>(0.126)</td>
<td>(0.366)</td>
<td>(0.247)</td>
<td>(0.465)</td>
<td>(0.128)</td>
<td>(0.217)</td>
<td>(0.431)</td>
</tr>
<tr>
<td>Unemployment-related</td>
<td>0.049(0.126)</td>
<td>0.150(0.126)</td>
<td>0.336(0.467)</td>
<td>0.097(0.208)</td>
<td>0.123(0.258)</td>
<td>0.013(0.451)</td>
<td>0.085(0.424)</td>
</tr>
<tr>
<td>insurance</td>
<td>(0.078)</td>
<td>(0.042)</td>
<td>(0.247)</td>
<td>(0.465)</td>
<td>(0.128)</td>
<td>(0.217)</td>
<td>(0.431)</td>
</tr>
<tr>
<td>Disability insurance</td>
<td>–0.1750(0.144)</td>
<td>–0.184(0.138)</td>
<td>0.280(0.231)</td>
<td>–0.093(0.084)</td>
<td>–0.449(0.202)</td>
<td>0.042(0.240)</td>
<td>–0.343(0.307)</td>
</tr>
<tr>
<td></td>
<td>(690)</td>
<td>(690)</td>
<td>(690)</td>
<td>(690)</td>
<td>(690)</td>
<td>(690)</td>
<td>(690)</td>
</tr>
<tr>
<td># Observations</td>
<td>37 939 858</td>
<td>25 554 728</td>
<td>1 280 145</td>
<td>3 791 759</td>
<td>4 469 054</td>
<td>1 919 349</td>
<td>688 913</td>
</tr>
</tbody>
</table>

Note: Each coefficient in this table is a result of a separate 2SLS regression based on Eqs. (3) and (4). Standard errors are clustered on commuting-zone. ‘*/’/*** indicates statistical significance at the 10/5/1 percent levels.

There is a causal influence on labor market outcomes three years before, so we expect the estimated effects to be zero in this case. As shown in Table E2, the placebo analysis displays no pattern of systematic “effects”. Two of the 26 estimated coefficients turn out to be statistically significant at the five percent level, but that is about what we can expect in the case of no systematic relationship.

7. Conclusion

The empirical evidence presented in this paper shows that there is a large grey area between social insurance programs targeted at unemployment-related and disability-related causes for insurance claims. Local labor demand conditions have a large and statistically significant influence on the caseload of temporary disability insurance programs, suggesting that temporary disability insurance in many cases is unemployment in disguise, in the sense that lack of realistic employment opportunities is the major cause behind the insurance claim. The effect of labor demand factors on the probability of entry into disability insurance is significant almost regardless of initial labor market state, and for new temporary disability entrants, the impact of labor demand conditions on the return-to-work probability is quite similar as it is for new unemployment insurance entrants.

Our findings indicate that there is a considerable element of judgement in relation to what kind of program a claimant is assigned to. As the two types of programs also entail different follow-up strategies, the choice of program is likely to have real consequences for future labor market outcomes. While the unemployment-related insurance programs typically contain activation requirements, in terms of monitored job search or active labor market program participation, the disability-related programs focus on pure income insurance and time for recovery.
If a claimant’s primary problem is joblessness, the assignment to a disability insurance program may be counterproductive, even when there are severe health problems involved. There is an increasing stock of empirical evidence showing that work is actually a healthy activity for workers with the illnesses and symptoms responsible for the vast majority of disability cases in industrialized countries (musculoskeletal diseases, back pain, and light mental disorders); see, e.g., Waddell (2004), Waddell and Burton (2006), OECD (2008, Chapter 4), van der Noordt et al. (2014), and Joyce et al. (2016). Hence, there is not only a blurred distinction between the two program types in terms of the primary causes of entry, but also in terms of the appropriate treatment and follow-up strategy. It may thus be time for a reconsideration of social insurance institutions that make a sharp distinction between unemployment and disability.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.labeco.2019.101767.

Appendix A. Industry composition in Norway

Fig. A1

Appendix B. Definition of monthly labor market states

To determine monthly labor market states for the whole adult population in Norway, we combine several administrative registers, covering demographics, earnings, business income, social insurance transfers, and education. Since our focus is on labor market outcomes, we restrict the population to individuals aged 18–61.

The four main states applied in this paper are constructed on the basis of a much more detailed state-space, comprising as much as 20 different states. Since it is possible to belong to several states in a given month, we apply a hierarchical ranking. The hierarchy is shown below, with lower numbers always “overwriting” higher numbers:

Hierarchy:

1. Passed away
2. Disability benefits, full-time (Uførepensjon)
3. Disability benefits, part-time, employed (Gradert uførepensjon)
4. Disability benefits, part-time, unemployed (Gradert uførepensjon)
5. Work assessment allowance, temporary disability (Arbeidsavklaringspenger, rehabiliteringspenger, atføringspenger)
6. Sick benefits (Sykepenger)
7. Parental benefits (Fødselspenger)
8. Unemployment benefits, full-time unemployed (Dagpenger, heltidssledge)

Fig. A1. Longitudinal and cross-sectional variation in industry-composition of employment in Norway

Note: Panel (a) shows a cross-plot of the number of employees in each industry in Norway in 1999 and 2016 on a log-scale. Panel (b) shows the same industries’ employment shares in 2016 (on the horizontal axis) and the corresponding coefficients of variation (across the 46 commuting zones). The symbols are used to indicate which main category the different industries belong to. Some data points are equipped with labels to give some flavor of the kind of level/type of aggregation.
9 Unemployment benefits, part-time unemployed (Daggengenber, 
delidssysselset)  
10 Employment scheme benefits (Tiltakspenger/venteotnand/individventnand)  
11 Social assistance (Sosialhjelp)  
12 Qualification benefits (Kvalifiersungstantnd)  
13 Enrolled in education  
14 Outside country  
15 Employed, full-time (> 30 h)  
16 Employed, part-time 1 (20–29 h)  
17 Employed, part-time 2 (4–19 h)  
18 Self-employed  
19 Transitional benefits for single parents (Overgangstantnd)  
20 No registered activity  

Based on this hierarchy, the four main states used in this paper are defined as follows:  

i) Employment: States 6, 7, 15, 16, 17, 18.  
ii) Disability-related social insurance: States 2, 3, 4, 5.  
iii) Unemployment-related social insurance: States 8, 9, 10, 11, 12.  

Appendix C: Ordinary Least Squares estimates  

Table C1 reports the OLS estimates from a regression of Eq. (3) with the actual employment rate (age 25–60) in the commuting zone in the year the outcome substituted for the predicted rate. These coefficients are thus directly comparable with the 2SLS estimates reported in Table 4.  

Appendix D: Alternative choices of outcome years  

Fig. D1 provides estimated labor demand effects (Eq. (4)) by initial base-year state and outcome-year (from 1 to 7 years after the base-year). Given that our main results (Table 4) consists of 28 different estimates, we have chosen to present the results for alternative choices of outcome years graphically, by plotting the alternative point estimates (with 95% confidence intervals) for each combination of sample and dependent variable in separate panels.  

Appendix E: Robustness  

Fig. E1 presents estimates based on the five alternative model specifications discussed in Section 6 above. Due to the large number of distinct estimates (see Table 4), we present the results graphically, and compare the alternative results for each of the 28 coefficients in separate panels. For convenience, the baseline results from Table 4 are repeated to the left in each panel.  

As explained in Section 6 above, a potential concern related to all the models based on Eqs. (3) and (4) is that the time variation in local industry weights within commuting zones may induce a simultaneity problem into the model, as these weights may be correlated to the error term in Eq. (3). In this appendix, we report estimates from a model where we keep industry weights constant across different base-years, and then include dummy variables for each combination of commuting zone and year of weight construction. This can only be done at the cost of making the instruments weaker, however, as the time distance between the construction of local weights and the predicted employment rates becomes larger. In order to use the data efficiently, it also implies that we have to reuse each observation several times, as the same outcomes can be examined with industry-weights constructed in different years. Let c be the year of local industry weight construction. We can then write the second stage equation as

\[
y_{ct} = a_{ct} + \varphi_{ct}y_{t-n} + \lambda e_{ct} + \beta e_{ct} + \varphi_{ct}y_{t-n} + \epsilon_{ct},
\]

where \((a_{ct}, \varphi_{ct})\) are separate commuting zone and year dummy variables for each year of weight-construction. The first stage equation is modified accordingly. The results from this alternative model is presented in Table E1. A comparison with the baseline results provided in Table 4 reveals that our main results are robust also with respect to this alternative model.  

Table E2 provides results from a placebo version of our model, where we have substituted outcomes observed three years before the base-year for the outcomes observed three years after. Although two of the 28 coefficients are statistically significant at the five percent level, we interpret these results as a confirmation of the absence of a systematic relationship.  

References  


Chapter 3:

Settlement neighborhood crime and
criminal behavior of young refugees in Norway
Chapter 4:

Material resources and well-being:
Evidence from an Ethiopian housing lottery
Material resources and well-being — Evidence from an Ethiopian housing lottery

Ashjørn G. Andersen, Andreas Kotsadam, Vincent Somville

A. Ragnar Frisch Centre for Economic Research, Norway
B. NHH Norwegian School of Economics, Norway
C. Michalm Institute, Norway
D. PROMENTA Research Center Department of Psychology, University of Oslo, Norway

ARTICLE INFO

Keywords:
- Mental Health
- Life Satisfaction
- Distress
- Wealth
- Housing
- Ethiopia

ABSTRACT

Do better material conditions improve well-being and mental health? Or does any positive relationship merely reflect that well-being promotes economic success? We compare winners and losers from a large Ethiopian housing lottery in a preregistered analysis. Winners gain access to better housing, experience a substantial increase in wealth, and report higher levels of overall life satisfaction and lower levels of financial distress. However, we find no average effects of winning on psychological distress. Our results suggest that not all aspects of well-being and mental health are equally sensitive to economic conditions.

1. Introduction

Ever since the United Nations included mental health and well-being among its Sustainable Development Goals, they have become a major policy concern internationally. Consequently, researchers increasingly emphasize the prevalence of common mental disorders (CMDs) and poor well-being in low- and middle-income countries and highlight poverty as both a cause and a consequence (Allouch, 2020; Deaton, 2008; Lund et al., 2010; Olesen et al., 2013; Patel et al., 2018; Ridley et al., 2020). However, the question remains as to the extent to which better material conditions reduce the prevalence of CMDs and improve well-being.

To inform this debate, we survey around 3000 winners and losers of an Ethiopian housing lottery two years after the draw. This lottery distributes purchase rights for new subsidized apartments to low- and middle-income households in Addis Ababa and is part of an ambitious urbanization program. Given that winning is random, we interpret the differences between winners and losers as the causal effect of winning the lottery. While winners gain access to better housing, they also experience a substantial increase in wealth through the ownership of real estate. According to our estimates, winners are on average 20 times wealthier than losers two years later.

* This research was financially supported through strategic funds from the Frisch Center and an NHH Norwegian School of Economics Småforsk grant. Kotsadam acknowledges support from the Norwegian Research Council (project number: 287766) “Field Experiments to Identify the Effects and Scope Conditions of Social Interactions”. Somville acknowledges support from the Norwegian Research Council (grant numbers: 250415 and 262675). The project was reviewed and approved by the Institutional Review Board (IRB) of the NHH Norwegian School of Economics (NHH-IRB 01/19). We thank Charlotte Hanlon and Markos Tesfaye for sharing their Amharic version of the Kessler psychological distress (K10) scale and the editor, Manisha Shah, as well as two anonymous reviewers who provided very valuable comments. We also thank Erik Ølolf Sørensen, Johannes Haushofer, and participants at the Essen Economics of Mental Health Workshop 2021, as well as participants at the Virtual Monthly Mental Health Economics Seminar Series (VMESS). A pre-analysis plan was submitted to the American Economic Association registry for Randomized Controlled Trials (No. AEARCTR-0003579) and all deviations from the plan are noted in the text.

* Corresponding author.

E-mail address: Vincent.Somville@nhh.no (V. Somville).

https://doi.org/10.1016/j.jhealeco.2022.102619
Received 26 April 2021; Received in revised form 3 April 2022; Accepted 3 April 2022
Available online 9 April 2022
0167-6296/© 2022 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/)
after the lottery. This scheme gives us an opportunity to study how winning the lottery and becoming substantially richer affects people’s mental health and well-being in a low-income context.

In addition to standard socioeconomic variables, we measure overall life satisfaction using standard questions from the World Values Survey, and psychological distress using the Kessler psychological distress scale (K10) (Kessler et al., 2002; 2003). We also include a set of survey questions to measure financial distress.

We find that winning the lottery increases overall life satisfaction by 0.2 standard deviations on average. This increase appears mostly driven by greater satisfaction with housing, neighborhood, and personal finances. Winners also report significantly lower levels of financial distress. More specifically, they are less likely to have inadequate means to cover household expenses, to have outstanding bills, and to have recently experienced financial difficulties more generally. However, we find no average effect of winning on psychological distress. In fact, the point estimate is remarkably close to zero and sufficiently precise that we can reject an effect of just 0.1 standard deviations.

These findings are robust to the inclusion of a large set of control variables, as well as the use of machine learning to select optimal controls. The results are also robust to different coding choices, and a bounds analysis — accounting for possible selective nonresponse — does not alter our main conclusions.

The lottery we study is clearly different from a randomized controlled trial where some individuals win cash and others do not. Through the lottery, winners simultaneously become wealthier and obtain access to better housing, and we cannot fully disentangle the effects of these changes. When we exploit the fact that only a minority of winners had relocated into the apartment won at the time of the interview, we find that the estimates for overall life satisfaction are very similar for both movers and nonmovers. We also find that both movers and nonmovers are more satisfied with their houses, neighborhoods, and financial situation. Of course, we must interpret this finding with caution given the risk of selection bias, but it nevertheless suggests that greater wealth rather than better housing drives our results.

In further trying to understand the mechanisms behind the effects, we first conduct a mediation analysis that shows that the share of the total effect on life satisfaction that is mediated by wealth is very high (range, 62–73 percent). The same analysis also shows that parts of the effects are mediated by financial distress, while there is no evidence of the effects being mediated by moving to the new property. We also test for heterogeneous effects of winning. One noteworthy finding is that poor individuals, but not the poorest, are more positively affected by winning the lottery both in terms of psychological distress and life satisfaction. This pattern may be explained by the fact that winners must make a down payment to receive their apartment. This upfront payment, as well as the monthly mortgage payments, may be associated with more distress for the poorest individuals. Despite experiencing lower financial distress on average, winners have higher debt and report having more difficulty in raising additional money for unexpected emergencies, which is likely because they have already borrowed to make the down payment.

The positive correlation between economic resources and life satisfaction and well-being is an almost universal finding (Clark, 2017; Deaton, 2008; Diener et al., 2010; Frijters et al., 2006; 2004; Haushofer and Fehr, 2014; Howell and Howell, 2008; Killingsworth, 2021). There is also increasing evidence of the negative association between poverty and mental health (Karimli et al., 2019; Ridley et al., 2020; Schilbach et al., 2016; Tampubolon and Hanandita, 2014). Nonetheless, existing evidence highlights that income is more strongly correlated with so-called evaluative measures of well-being, such as life satisfaction, than with more affective measures, such as questions about the frequency of various positive or negative feelings (Kahneman and Deaton, 2010). In analyzing the effects on different subindices of life satisfaction, we see that the positive effects are concentrated in the financial and physical environment (neighborhood, home, and leisure) domains. Hence, the difference in results across life satisfaction and psychological distress in our study may not be due to a difference between evaluative and affective measures but rather to a difference between satisfaction with financial status (and other material dimensions) and more intangible dimensions, including psychological distress.

Our findings provide causal evidence on the relationship between material conditions and mental health and well-being. Nevertheless, this is not the first analysis to move beyond descriptive correlations to make causal claims. Likewise, other studies have exploited variations in economic resources from natural experiments. For example, using tax rebates, Lachowska (2017) finds that increased income reduces stress and worry in the US. Also in the US, Schwandt (2018) employs stock price fluctuations and finds that increases in wealth improve mental health. There is also evidence of mental health effects from variations in income from casinos among Native Americans (Costello et al., 2003; 2010; Wolfe et al., 2012).

Other studies have also used lotteries to investigate the effects of monetary gains on well-being and mental health. The best evidence from lotteries to date is from Sweden, where Lindqvist et al. (2020) compare winners with equal probabilities of winning in a large sample using a prereregistered analysis. They find a persistent positive relationship between the lottery amount won and overall life satisfaction. However, they similarly show no significant effects on mental health. Gardner and Oswald (2007) and Apouey and Clark (2015) in the UK and Lindahl (2005) in Sweden find that large lottery wins lead to improvements in mental health. However, these studies compare winners from different lotteries and lack information about how often people played. It is therefore unclear whether the winners of different amounts are drawn from the same populations. The sample sizes in these studies are also small (ranging from just 137 to 674 winners). In contrast, Kuhn et al. (2011) find no effect on happiness of winning in a

---

1 We use the Amharic version of the K10, tested and used in Ethiopia by Fekadu et al. (2014); Tesfaye et al. (2010, 2016), and which Charlotte Hanlon and Markos Tesfaye graciously shared with us.

2 Better housing and neighborhood quality have been repeatedly identified as associated with a lower prevalence of CMDs and better well-being (Abas and Broadhead, 1997; Alloush and Bloom; Amoran et al., 2005; Cattaneo et al., 2009; Danaci et al., 2002; Gureje et al., 2007; Kim et al., 2002; Ludvig et al., 2012; Lund et al., 2010; Patel et al., 2006; 1998; Sabin et al., 2003).
Dutch lottery where they were able to compare 223 winners and 477 losers in the same lottery, even though they were unable to reject large effects. We are also aware of an unpublished working paper using data from an earlier Ethiopian housing lottery, in which Franklin (2019) investigates the demand for state subsidized housing in Ethiopia. In the Appendix, the author also reports exploratory results for well-being and mental health among lottery participants (in Appendix Table A.26). He finds that winning reduces a composite index of anxiety and depression among winners, but the effect of -0.11 is only statistically significant at the 10 percent level.  

Given the context of our study, we also contribute to the research on the relationship between economic resources and mental health in low-income countries, where most of the causal evidence comes from cash transfer programs. In a recent meta-analysis of 38 cash transfer studies covering the period 2000–2020, McGuire et al. (2020) find a positive effect of 0.1 standard deviations on a composite index of mental health and well-being, whereas the effect is smaller for mental health in isolation. The fact that the main source of heterogeneity in the effects is the size of the transfer highlights the need for studies of more radical changes in economic conditions, such as those presented here. Ridley et al. (2020) focus on mental health and also include poverty-alleviating programs other than cash transfers. As in McGuire et al. (2020), they find an overall positive effect of about 0.1 standard deviations. Romero et al. (2021) conduct a systematic review of 57 studies that investigate the effect of economic interventions on well-being. They also find that the average effect is 0.1 standard deviations. Interestingly, they include an analysis of 10 housing voucher programs, all of which are from the US and nine of them study moving-to-opportunity programs. These programs have a meta-analytic effect on mental health outcomes of 0.07 standard deviations. Finally, Zimmerman et al. (2021) conduct a meta-analysis of the effects of cash transfers on the mental health of children and young adults. Their target population is different from ours, and from the two earlier reviews, but their focus on mental health is relevant to this study. They find substantial heterogeneity between studies and conclude that cash transfers have no impact on depressive symptoms.

The present study differs from this literature along several dimensions. As mentioned, the lottery winners in our sample see an exceptionally large increase in wealth, which is presumably permanent and relatively certain given the stability of the real estate market in a fast-growing city such as Addis Ababa. This contrasts with the relatively small short-term income changes induced by temporary cash transfers. As well-being and mental health are influenced by uncertainty and worries (Ridley et al., 2020), a permanent increase in wealth could exert even stronger effects on well-being and mental health than could temporary transfers. The fact that we observe a reduction in financial distress among winners also suggests that we could expect a beneficial effect from fewer worries. In light of this, our null result on psychological distress (and the nonfinancial and physical environment aspects of life satisfaction) is quite stark. One important difference as compared with the cash transfer literature is that our lottery induced people to borrow and make down payments, which may have offsetting effects for the poorest individuals in our sample. The fact that we find larger effects on life satisfaction, as well as significant improvements in psychological distress, for the poor, yet not the poorest, is consistent with such an interpretation.

The remainder of the paper is structured as follows. We describe the lottery and the context in Section 2 and present the data in Section 3. We describe our empirical strategy and discuss the main results in Section 4. We discuss plausible mechanisms and conduct additional exploratory analyses in Section 5. We summarize the study and provide concluding remarks in Section 6.

2. The lottery

The housing lottery we consider is part of a large-scale urbanization policy known as the Integrated Housing and Development Programme (IHDP), which aims to facilitate access to quality housing for low- and middle-income groups in Addis Ababa, Ethiopia. The apartments are sold at highly subsidized prices and home-buyers are given access to finance through the Commercial Bank of Ethiopia (CBE). Given the excess demand for housing at subsidized prices, condominium apartments are distributed through a computer-based lottery among eligible applicants.

Eligibility for the lottery is based on three requirements: (i) having resided in Addis Ababa for at least the previous two years, (ii) not having any other house or lease land registered (in one’s own or a spouse’s name), and (iii) having opened a savings account at the CBE and deposited the required monthly savings for at least 29 months (with no breaks in saving longer than six months).

The IHDP is a large-scale and comprehensive program. During the initial registration in 2005, more than 300,000 households in Addis Ababa signed up for the program, corresponding to roughly half of the city’s population. When registering for the program, applicants must select the desired apartment type (studio, one-, two-, or three-bedroom apartments). As supply and demand vary by unit type, separate lotteries are held for each type of apartment. Within each lottery, quotas exist for women, the disabled, and government employees. First, 30 percent of the winners are drawn from among female applicants. Then 20 percent of the winners

---

5 Note that this analysis was not included in the 2018 version of the paper and was not one of the main preregistered outcomes. In addition, the statistical power to detect an effect of 0.1 is only 0.47 given the sample size in that study.

6 Studies that assess the effects of cash transfers on psychological well-being and mental health include Alzua et al. (2019); Angeles et al. (2019); Baird et al. (2013); Bando et al. (2020); Blattman et al. (2020, 2017); Chen et al. (2019); Egger et al. (2019); Galama et al. (2017); Galiani et al. (2016); Han and Gao (2020); Haushofer et al. (2020a,b); Haushofer and Shapiro (2016, 2018); Heath et al. (2020); Hjelm et al. (2017); Hussam et al. (2021); Klümborn et al. (2018, 2019, 2016); Macours et al. (2012); Ohrnberger et al. (2020a,b); Ozer et al. (2011); Paxson and Schady (2010); Salinas-Rodríguez et al. (2014); Schatz et al. (2012). Rather than discussing all of these, we refer the interested reader to recent reviews by McGuire et al. (2020), Ridley et al. (2020), Romero et al. (2021), and Zimmerman et al. (2021).

7 Their meta-analysis includes 12 cash transfers and six multifaceted antipoverty programs.

8 Our study is also designed to detect an effect of 0.1 standard deviations with a power of 0.8 at the 0.05 level of significance.
are drawn from among government employees. Finally, there is a five percent quota for those with physical disabilities. All quotas are decided upon after registration but before the lottery draw.

Currently, there have been two rounds of registrations and 13 lotteries. We focus on the first round of registration and the 11th lottery rounds, which took place in 2016.7

Lottery winners must pay at least 20 percent of the apartment price up-front and are offered access to finance for the remaining 80 percent through the CBE. Given this payment scheme, the program has been labeled the “20/80 program.”

The 11th round of the lottery distributed the purchase rights for 12,027 apartments (excluding three-bedroom units).8 Only individuals who had registered in 2005 were included in the draw. Upon winning the lottery, prospective homeowners were required to make the 20 percent down payment before they could sign the contract and receive the keys to their apartment. Around 95 percent of the winners initially drawn were able to do this. They are then free to rent out their apartment but are not allowed to sell it within the first five years.9 At the time of the survey, two years after the lottery, 30 percent of the winners had moved into their apartment, 31 percent were renting it out, 32 percent were currently empty, but with the owner planning to move in (21 percent) or rent it out (11 percent), and in two percent of the cases, the apartment was used on a rent-free basis by relatives.

3. Data

We designed and collected survey data for the lottery winners and losers in collaboration with the Ethiopian Development Research Institute (EDRI). We sampled applicants who registered in the first round (in 2005) for a studio, a one-, or a two-bedroom apartment, and who were eligible for the 11th lottery in 2016. As noted, there were special quotas for women, government employees, and people with physical disabilities. Therefore, we needed to obtain information on these variables.

There are two different administrative lists pertaining to the lottery: one for winners and one for losers. Therefore, EDRI obtained two types of lists from the Addis Ababa Housing Development and Administration Agency: one for winners and one for losers. Starting with the winners, we randomly sampled 2200 individuals on this list who had unique telephone numbers and who had not won a three-bedroom apartment. For this ‘winners’ sample,” we have information about the apartment type, gender, and public sector employment at the time of the registration. We also have information about the location of the apartments won. We did not have information about physical disability status at registration, so we had to ask them about this separately during the survey.

EDRI also obtained the list of individuals who registered in 2005 and qualified for the 11th lottery, but who did not win it (and did not win the 12th lottery either, which was held between the 11th lottery and our survey). This list includes information about the type of apartment the individuals applied for and about physical disability status. We obtained employment status and gender data during the survey.10 We also ranked all individuals on this list randomly and then selected a random sample of 2200 losers (stratified by gender within each apartment type). This is our “losers’ sample.” We then aggregated the winners’ and losers’ samples and randomized the order again. We created a new ID variable and keep only the people’s IDs, names, and phone numbers before sending the list to the data collection team. In this way, the individual’s status (i.e., winner or loser) is blinded for the enumerators, and we avoid issues with confounding factors because of different timing and different enumerators. EDRI interviewed the sampled individuals by phone using the survey questionnaire developed by the research team. The survey took around 20 minutes to complete, and the respondents were given ETB 50 in compensation. EDRI was told to stop after around 3000 completed interviews. The survey respondents were paid with mobile money directly after the interview was conducted.

3.1. Survey measures

Our first two outcome variables measure psychological well-being. In addition, we examine the effects on financial distress and also collect data on features that serve as control variables. Next, we describe these variables.11

Our first two outcomes are overall life satisfaction and psychological distress. For overall life satisfaction, we use the standard measure from the World Values Survey, which asks respondents: “Using a scale from 0 to 10, where 0 means ‘not at all satisfied,’ and 10 is ‘completely satisfied,’ how satisfied are you with your life as a whole these days?” We standardize the responses by subtracting the mean and dividing by the standard deviation (both from the losers group).

To further explore this dimension, we also include measures of domain-specific satisfaction. In particular, we question respondents about how satisfied they are with their health, leisure time, financial situation, friends, relatives, home, neighborhood, work, and with Ethiopian society. Responses are given on a scale from 0 to 5, where 0 is very dissatisfied and 5 is very satisfied. These variables are standardized in the same manner as the responses to the overall life satisfaction question.12

---

7 The 12th round, conducted in 2018, was unusually small with only 2607 apartments and the 13th round took place in March 2019 after data collection for the project was completed.
8 We excluded applicants for three-bedroom apartments because almost everyone in this group had received an apartment at the time of sampling.
9 A small share (four percent) of the winners in our sample managed to sell their apartment, despite these rules.
10 We first inferred the individual’s gender from their first name and later confirmed it during the interview.
11 The full survey is available in Appendix Section.
12 In the pre-analysis plan, we stated that we would dichotomize each variable by choosing the cut-off that would divide the losers group into two groups of as equal sizes as possible. We included these results in Appendix Section to show that it makes no qualitative difference to our estimates.
We measure psychological distress using the Kessler psychological distress scale (K10) (Kessler et al., 2002; 2003). The K10 scale contains 10 questions concerning experienced symptoms of depression and anxiety in the past 30 days. Respondents are asked how often they have felt:

(i) ...tired out for no good reason;
(ii) ...nervous;
(iii) ...so nervous that nothing could calm them down;
(iv) ...hopeless;
(v) ...restless or fidgety;
(vi) ...so restless they could not sit still;
(vii) ...depressed;
(viii) ...that everything was an effort;
(ix) ...so sad that nothing could cheer them up;
(x) ...worthless.

Responses are given on a five-point scale ranging from none of the time to all of the time. The range of scores is between 10 and 50, where higher scores indicate higher distress (Andersen et al., 2011; Andrews and Slade, 2001). The K10 scale is used widely, including in the World Mental Health Survey, and has been translated and validated in many different contexts, including in Ethiopia (Fekadu et al., 2014; Tesfaye et al., 2010; 2016). The K10 scale is highly correlated with other screening scales for CMDs (Patel et al., 2008) and has the advantage of being short and concise. The internal consistency of the index is high. We obtain a Cronbach’s alpha of 0.9 using our data, which exactly matches the value reported by (Tesfaye et al., 2010). For comparability, we standardize the overall K10 score in the same way as for the life satisfaction question. To explore various aspects of distress, we also report effect estimates for the individual items on the scale (also standardized).

In the literature, it is common to use cut-off scores to separate the levels of distress. Suggested score categories are: 10–19 (individual is likely well), 20–24 (indicating mild mental disorder), 25–29 (indicating moderate mental disorder), and 30–50 (indicating severe mental disorder). According to these cut-off scores, 78 percent of the losers show no signs of mental disorder, 14 percent have mild mental disorders, and eight percent suffer from moderate to severe mental disorders. The literature has emphasized that women bear a disproportionate share of the burden of mental illness (James et al., 2018), and this is also the case in our sample, where the shares of women falling into each category of mental disorder are 73, 15, and 12 percent, whereas the corresponding shares for men are 81, 13, and six percent, respectively. We did not prespecify the use of cut-off scores, but in Appendix Section, we show that our conclusions are the same when applying thresholds. We also show that alternatively using the Kessler K6 scale — nested in the K10 scale but including only six of the above 10 items — does not affect our conclusions either.

To assess the effect of winning the lottery on economic resources, we measure the wealth and experienced financial distress of respondents. Based on the reported asset values (including real estate) and liabilities, with all currency values in Ethiopian birr (ETB), we calculate their housing-related wealth and net wealth. At the time of the survey in 2018, ETB 1000 was equivalent to around USD 36. We also asked respondents about whether they were richer today than five years ago, whether they expected to be richer five years from now, and whether they perceived themselves as richer, equally rich, or poorer than other Ethiopians. In addition, we constructed an asset index based on whether the households owned a radio, TV, refrigerator, car, computer, tablet, satellite dish, smartphone, or an electric mitad (a common cooking appliance like a grill in Ethiopia).

Finally, because economic distress may be an important channel through which economic circumstances affect well-being, we include four commonly used measures of financial distress. We first ask “If you suddenly ended up in an unforeseen situation, where you must raise ETB 20,000, would you be able to?” We code the response as a binary indicator equal to one if the answer is no. We then ask three questions about the economic situation of each respondent’s family during the last six months. Specifically, we ask whether they have had inadequate money to cope with family expenses (never, rarely, sometimes, or always), if they have delayed the payment of bills because of financial difficulty (never, rarely, sometimes, or always), and what the economic condition of the family has been like (no, some, considerable, or much financial difficulty).

For comparability with our main outcomes, we standardize each of the items relating to financial distress by subtracting the mean and dividing by the standard deviation of the losers group. We then construct a financial distress index by adding the four standardized items together and standardizing the sum in the same way. The four items are highly correlated, and the internal consistency of the index, as measured by Cronbach’s alpha, is 0.81. We present the effect of winning on this financial distress index along with our main results because it is seen as a key channel for the effects of economic resources on distress.

We also asked the respondents about demographic variables (ethnicity, religion, age, region of birth, and partner status) and earnings just before the lottery (in 2015) and at registration (in 2005) for themselves and their partners. The coding of the control variables is described in the Appendix Table.

---

13 We employ an Amharic version shared by Hanlon and Tesfaye.
14 The values for these variables are missing for about 40 and 60 percent of respondents because of missing or inconsistent information on one or more of the variables, respectively. As specified in the pre-analysis plan, we calculate the bounds on the effect of winning the lottery on wealth. Appendix Table shows that the difference in wealth between lottery winners and losers is still large and significantly different from zero, even if we make very extreme assumptions about the values of the missing observations.
15 In the pre-analysis plan, we stated that we would dichotomize each of the financial distress items and we show in Appendix Section that this makes no qualitative difference to our main results. We chose to present a standardized index in the main paper to ease comparisons across outcomes.
Table 1
Descriptive statistics.

<table>
<thead>
<tr>
<th></th>
<th>Total</th>
<th>Winner</th>
<th>Loser</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>Winner</td>
<td>0.49</td>
<td>(0.50)</td>
<td>1.00</td>
</tr>
<tr>
<td>Strata variables</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>0.42</td>
<td>(0.49)</td>
<td>0.45</td>
</tr>
<tr>
<td>Government</td>
<td>0.22</td>
<td>(0.41)</td>
<td>0.30</td>
</tr>
<tr>
<td>Disabled</td>
<td>0.03</td>
<td>(0.17)</td>
<td>0.06</td>
</tr>
<tr>
<td>Studio</td>
<td>0.20</td>
<td>(0.40)</td>
<td>0.20</td>
</tr>
<tr>
<td>One bedroom</td>
<td>0.54</td>
<td>(0.50)</td>
<td>0.53</td>
</tr>
<tr>
<td>Two bedroom</td>
<td>0.26</td>
<td>(0.44)</td>
<td>0.26</td>
</tr>
<tr>
<td>Other control</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>variables</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>42.81</td>
<td>(9.60)</td>
<td>43.38</td>
</tr>
<tr>
<td>Orthodox</td>
<td>0.76</td>
<td>(0.43)</td>
<td>0.77</td>
</tr>
<tr>
<td>Muslim</td>
<td>0.11</td>
<td>(0.32)</td>
<td>0.09</td>
</tr>
<tr>
<td>Protestant</td>
<td>0.12</td>
<td>(0.32)</td>
<td>0.12</td>
</tr>
<tr>
<td>Amhara</td>
<td>0.37</td>
<td>(0.48)</td>
<td>0.38</td>
</tr>
<tr>
<td>Gurage</td>
<td>0.17</td>
<td>(0.37)</td>
<td>0.15</td>
</tr>
<tr>
<td>Oromo</td>
<td>0.16</td>
<td>(0.37)</td>
<td>0.16</td>
</tr>
<tr>
<td>Tigray</td>
<td>0.08</td>
<td>(0.28)</td>
<td>0.09</td>
</tr>
<tr>
<td>Born in Addis</td>
<td>0.45</td>
<td>(0.50)</td>
<td>0.42</td>
</tr>
<tr>
<td>Ababa</td>
<td>0.18</td>
<td>(0.38)</td>
<td>0.19</td>
</tr>
<tr>
<td>Born in Oromia</td>
<td>0.15</td>
<td>(0.36)</td>
<td>0.16</td>
</tr>
<tr>
<td>Born in SNNP</td>
<td>0.14</td>
<td>(0.35)</td>
<td>0.14</td>
</tr>
<tr>
<td>Born in Tigray</td>
<td>0.06</td>
<td>(0.24)</td>
<td>0.08</td>
</tr>
<tr>
<td>Earnings 2005</td>
<td>5.13</td>
<td>(3.19)</td>
<td>5.22</td>
</tr>
<tr>
<td>(at reg.)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Earnings 2015</td>
<td>7.05</td>
<td>(3.03)</td>
<td>7.14</td>
</tr>
<tr>
<td>(at reg.)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Partner earnings</td>
<td>0.92</td>
<td>(2.47)</td>
<td>0.92</td>
</tr>
<tr>
<td>2005 (at reg.)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Partner earnings</td>
<td>1.57</td>
<td>(3.25)</td>
<td>1.61</td>
</tr>
<tr>
<td>2015 (at reg.)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Partner 2015</td>
<td>0.32</td>
<td>(0.46)</td>
<td>0.31</td>
</tr>
<tr>
<td>N</td>
<td>3049</td>
<td>1485</td>
<td>1564</td>
</tr>
</tbody>
</table>

Note: The table shows the means and standard deviations of individual characteristics over the whole sample and separately among the lottery winners and losers. An F-test of whether all “Other control variables” jointly predict winning after the strata variables are controlled for returned a value of 0.28 (p = .60).

3.2. Descriptive statistics and balance test

In this section, we describe the sample across some important dimensions and check whether we can identify any noteworthy differences between winners and losers prior to the draw. Table 1 presents descriptive statistics for all individuals and for the winners and losers separately. We can see that 49 percent of the final sample are winners. Regarding the strata variables, 42 percent of the respondents are women, while the shares registered for a studio and a one- or two-bedroom apartment are 20, 54, and 26 percent, respectively.

Although we stratified the sample by gender, the share of women is slightly higher in the winner group (45 vs. 40 percent). This is because the gender inferred from respondent names is not always accurate and the gender was updated during the interview.16 As expected, given the quotas for these groups, the shares of government employees and those with physical disabilities are higher among the winners (30 and 5.8 percent, respectively) than among the losers (14 and 0.3 percent, respectively). Given that this information was not available beforehand, we could not stratify the sampling on these variables. We describe these issues in detail in Appendix Section, where we also show that alternative coding choices have little consequence for the main results.

The mean age of respondents is around 43 years (which implies that they were on average 29–30 years when they signed up in 2005), and the most common religions are Orthodox Christianity (76 percent), Protestantism (12 percent), and Islam (11 percent).17 The most common ethnic groups are Amhara (37 percent), Gurage (17 percent), Oromo (16 percent), and Tigray (eight percent), while the most common regions of birth are Addis Ababa (45 percent), Amhara (18 percent), Oromia (15 percent), Southern Nations, Nationalities, and People’s Region (SNNP) (14 percent), and Tigray (six percent).

We test for balance in the control variables across the winner and loser groups by regressing the “winner” variable on the control variables described, while controlling for the strata fixed effects $X_i$ (gender, government employee, disabled, and apartment type). Based on the F-test (see note below Table 1), we reject the hypothesis that these variables jointly predict winning. In Table, we also present regressions of the treatment on each variable individually and together, while controlling for the strata variables. While the

---

16 If we instead use the gender variable based on the names (as we did for the sampling), we find that the shares are similar for both groups (44–45 percent).

17 For all categorical variables, we pool small groups accounting for less than five percent of the population.
F-test shows that there is balance in general, there are differences between the winners and losers on some variables. As explained in the following section, we therefore also present our results where we control for all control variables as well as for a subset of variables selected through a double robust LASSO procedure.

3.3. Attrition and nonresponse

EDRI was given a list of 4400 individuals in total; however, 1082 of the telephone numbers were invalid. There was no difference between winners and losers in the probability of having an invalid number. In total, EDRI called 3318 people and completed interviews with 3049 individuals (1,485 winners and 1564 losers). Therefore, the response rate is 92 percent. As seen in Table, the share of people declining to be interviewed (unwilling) is significantly larger among the winners. There is no difference between winners and losers in the share of people who moved abroad, were never available to answer the survey, had passed away, or for which the person answering said it was a wrong number.

The total response rate is statistically significantly different between winners and losers after controlling for gender and apartment type (the only strata variables available for both winners and losers not answering the survey). In Appendix Section A2a, we present the results from a pre-specified bounds analysis and show that our main results are robust to reasonable assumptions about the potential values of the missing observations.

We also have missing values for some of the covariates. Among the prespecified controls, we have complete information about all but four: earnings in 2005, earnings in 2015, partner earnings in 2005, and partner earnings in 2015. These are recall questions and some respondents could not remember the answers. In the Appendix Table, we show that these missing values are not correlated with winning the lottery and that the lottery effect is not different for people with and without missing values. When we use these additional covariates in a regression, we use the missing-indicator method (White and Thompson, 2005): we add an indicator variable equal to one to the regression when the variable is missing and set the missing values to zero.19

4. Empirical strategy and main results

To test the effects of winning the lottery on individual i’s outcomes, we regress the outcome of interest \( Y_i \) on \( T_i \), a dummy variable equal to one if the individual has won the lottery, while controlling for the set of strata covariates \( S_i \) (gender, public sector employment, disability, and apartment type):

\[
Y_i = \beta T_i + \theta S_i + \epsilon_i
\]

This is our main specification as described in the pre-analysis plan.

We show that the results are robust to including the full set of control variables, as well as to a subset of control variables selected using the post double LASSO approach of Belloni et al. (2014).20 Because the treatment is randomized at the individual level, we use robust standard errors without any clustering.

4.1. Effects of winning on wealth and disposable income

As noted, we interpret the effects of winning the lottery primarily in terms of a wealth effect. To substantiate this interpretation, we begin by showing the effect of winning on wealth.

We define net wealth as the total reported value of any real estate owned plus savings minus debt of any sort. According to this measure, lottery winners are clearly wealthier than losers. At the time of the interview (two years after the lottery), the average net wealth reported by winners is around ETB 450,000 (roughly USD 45,000 PPP adjusted). This is more than 20 times the amount reported by losers (ETB 20,000 or less than USD 2000 PPP adjusted), and the difference corresponds to around 15 years of average earnings in our data. The net wealth distribution for the two groups is illustrated in Fig. 1.

The main weakness of our wealth measures is missing values for part of the sample. This is because some respondents were unable to supply a specific market value for their real estate, and because some did not report their wealth during the interview. In Table 2, we report bounded estimates for the lottery effect on wealth. We obtain the lower bounds of the lottery effect by replacing missing observations among the winners (losers) by that group’s mean value minus (plus) 0.20 standard deviations of the loser group. The upper bounds of the effects are constructed in a symmetrical fashion (also see Appendix 3.2). This exercise shows that winning reduces savings by ETB 4194 to ETB 12,800 (column 1) and increases debt by ETB 120,834 to 128,939 (column 2). However, this decrease in net savings is more than offset by the increase in housing wealth (defined as the respondent’s expected selling price of any housing

---

18 This is unsurprising because the lottery participants registered in 2005, i.e., 13 years prior to the data collection. However, the fact that phone numbers on the participants lists are outdated does not imply that some of the winners miss out. Shortly after the lottery draws, which are subject to intense media coverage, the list of winners is published (both in print and online), so that winners can themselves contact the authorities to claim their apartment.

19 This method can lead to biased estimates in some cases (Jones, 1996). Fortunately, our estimates remain very stable whether we use this imputation method, only use the subsample with complete data, or do the estimations without controls.

20 To the extent there is concern about imbalance, the LASSO selection approach is also helpful as it precisely selects those variables that are correlated with both the treatment and the outcomes.
units owned), which increases by ETB 563,533 to ETB 588,187 (column 3). As a consequence, net wealth increases significantly by ETB 395,010 to ETB 456,794 (column 4).

Turning to the more qualitative aspects, Table 2 shows that winners also perceive themselves to be richer than five years ago (the estimated effect is 6.5 percentage points relative to a mean of 71 percent among the losers) and expect to become even richer over the next five years (by 1.4 percentage points). Finally, a larger share of winners perceives themselves to be as rich as or richer than Ethiopians in general (by 10 percentage points). This analysis suggests that winning the lottery has a substantial impact on self-assessed wealth and economic position.

We find no effects on household assets, perhaps because such an effect takes longer to materialize. Another explanation may be that winners have not invested in household assets because they spend a large share of their income on mortgage payments, and that their disposable income is almost unaffected by winning (at least in the short run). This is confirmed by Tables and in the Appendix, which show that winning increases both household expenditures and income.

To summarize, while winners gain the ownership of a house, they will often need to borrow money to finance the down payment as well as the mortgage payments. The economic impact of winning is therefore a massive increase in wealth but also reduced savings and increased debt. However, the net wealth effect of winning the lottery is substantial, corresponding to 15 years of average earnings, and winners are 20 times wealthier than losers on average. Of course, winners realize this disparity and are more likely than losers to report being wealthier than five years ago and being wealthier than other Ethiopians generally.

Table 2

<table>
<thead>
<tr>
<th>Winner</th>
<th>Total savings (1)</th>
<th>Total debt (2)</th>
<th>Housing wealth (3)</th>
<th>Net wealth (4)</th>
<th>Richer than 5 years ago (5)</th>
<th>Richer in 5 years (6)</th>
<th>Perceived position (7)</th>
<th>Asset index (8)</th>
</tr>
</thead>
</table>
| Lower bound | -12.800***  
(1.045) | 120.834***  
(4.822) | 563.533***  
(8.133) | 395.010***  
(7.245) | 0.065***  
(0.016) | 0.014*  
(0.008) | 0.104***  
(0.017) | 0.046  
(0.034) |
| Upper bound | -4.194***  
(1.047) | 128.939***  
(4.826) | 588.187***  
(8.149) | 456.794***  
(7.279) | 0.065***  
(0.016) | 0.014*  
(0.008) | 0.104***  
(0.017) | 0.046  
(0.034) |
| Mean (losers) | 18.014  
3049 | 7.329  
3049 | 6.859  
3049 | 20.407  
3049 | 0.706  
3049 | 0.941  
3049 | 0.634  
3049 | 0.000  
3049 |

Note: The table reports the estimated effects of winning the lottery. Robust standard errors are in parentheses. We control for the stratification variables in all estimations. P-values are ≤ 0.01***, ≤ 0.05**, and ≤ 0.1*. Wealth-related variables in (1)–(4) are in ETB 1000 (ETB 1000 was equivalent to around USD 36 at the time of the survey, in 2018).
Table 3
Correlates of well-being.

<table>
<thead>
<tr>
<th></th>
<th>Overall life satisfaction</th>
<th>Financial distress</th>
<th>Psychological distress</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>Female</td>
<td>-0.170***</td>
<td>-0.145**</td>
<td>0.289***</td>
</tr>
<tr>
<td></td>
<td>(0.051)</td>
<td>(0.062)</td>
<td>(0.050)</td>
</tr>
<tr>
<td>Public employee</td>
<td>0.046</td>
<td>0.025</td>
<td>-0.054</td>
</tr>
<tr>
<td></td>
<td>(0.068)</td>
<td>(0.070)</td>
<td>(0.072)</td>
</tr>
<tr>
<td>Disabled</td>
<td>-1.477***</td>
<td>-1.495***</td>
<td>0.703***</td>
</tr>
<tr>
<td></td>
<td>(0.260)</td>
<td>(0.207)</td>
<td>(0.317)</td>
</tr>
<tr>
<td>One bedroom</td>
<td>0.250***</td>
<td>0.177***</td>
<td>-0.497***</td>
</tr>
<tr>
<td></td>
<td>(0.068)</td>
<td>(0.068)</td>
<td>(0.063)</td>
</tr>
<tr>
<td>Two bedroom</td>
<td>0.407***</td>
<td>0.307***</td>
<td>-0.851**</td>
</tr>
<tr>
<td></td>
<td>(0.077)</td>
<td>(0.078)</td>
<td>(0.073)</td>
</tr>
<tr>
<td>Age</td>
<td>-0.004</td>
<td>0.004</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>Orthodox</td>
<td>-0.510***</td>
<td>0.201</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.190)</td>
<td>(0.146)</td>
<td></td>
</tr>
<tr>
<td>Muslim</td>
<td>-0.540***</td>
<td>0.321**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.202)</td>
<td>(0.163)</td>
<td></td>
</tr>
<tr>
<td>Protestant</td>
<td>-0.188</td>
<td>0.128</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.203)</td>
<td>(0.162)</td>
<td></td>
</tr>
<tr>
<td>Amhara</td>
<td>0.039</td>
<td>0.011</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.078)</td>
<td>(0.073)</td>
<td></td>
</tr>
<tr>
<td>Gurage</td>
<td>-0.257***</td>
<td>0.163**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.086)</td>
<td>(0.083)</td>
<td></td>
</tr>
<tr>
<td>Oromo</td>
<td>0.086</td>
<td>0.010</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.088)</td>
<td>(0.079)</td>
<td></td>
</tr>
<tr>
<td>Tigray</td>
<td>0.127</td>
<td>0.044</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.155)</td>
<td>(0.146)</td>
<td></td>
</tr>
<tr>
<td>Born in Addis</td>
<td>0.302*</td>
<td>-0.227</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.166)</td>
<td>(0.174)</td>
<td></td>
</tr>
<tr>
<td>Born in Amhara</td>
<td>0.077</td>
<td>-0.006</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.175)</td>
<td>(0.184)</td>
<td></td>
</tr>
<tr>
<td>Born in Oromia</td>
<td>0.056</td>
<td>-0.061</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.177)</td>
<td>(0.187)</td>
<td></td>
</tr>
<tr>
<td>Born in SNNP</td>
<td>0.245</td>
<td>0.029</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.187)</td>
<td>(0.195)</td>
<td></td>
</tr>
<tr>
<td>Born in Tigray</td>
<td>0.289</td>
<td>-0.230</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.205)</td>
<td>(0.227)</td>
<td></td>
</tr>
<tr>
<td>Earnings 2005 (at reg.)</td>
<td>-0.014</td>
<td>0.010</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td>(0.009)</td>
<td></td>
</tr>
<tr>
<td>Earnings 2015</td>
<td>0.032***</td>
<td>-0.053***</td>
<td>-0.033***</td>
</tr>
<tr>
<td></td>
<td>(0.010)</td>
<td>(0.010)</td>
<td></td>
</tr>
<tr>
<td>Partner earnings 2005 (at reg.)</td>
<td>-0.009</td>
<td>0.003</td>
<td>0.011</td>
</tr>
<tr>
<td></td>
<td>(0.015)</td>
<td>(0.014)</td>
<td></td>
</tr>
<tr>
<td>Partner earnings 2015</td>
<td>0.007</td>
<td>-0.037***</td>
<td>-0.008</td>
</tr>
<tr>
<td></td>
<td>(0.011)</td>
<td>(0.010)</td>
<td></td>
</tr>
<tr>
<td>Partner 2005 (at reg.)</td>
<td>-0.095</td>
<td>0.118</td>
<td>0.034</td>
</tr>
<tr>
<td></td>
<td>(0.084)</td>
<td>(0.075)</td>
<td></td>
</tr>
<tr>
<td>Partner 2015</td>
<td>0.087</td>
<td>0.164**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.072)</td>
<td>(0.066)</td>
<td></td>
</tr>
</tbody>
</table>

Note: The table shows ordinary least squares (OLS) estimates for the correlation between the main outcomes and baseline characteristics among the losers. Robust standard errors are in parentheses. P-values are ≤ 0.01***, ≤ 0.05**, and ≤ 0.1*. The dependent variables are standardized (a mean of zero and a standard deviation of one).

4.2. Main results

Having shown that treatment status appears to be randomly assigned conditional on the strata and that there is a substantial effect of winning on wealth, we now turn to our well-being outcomes. In addition to the prespecified primary outcomes of life satisfaction and psychological distress, we shall also present results on financial distress.

To obtain a first impression of the general correlates of well-being, Table 3 details how the outcomes correlate with the strata and other control variables among the lottery losers. As shown in column 1, overall life satisfaction tends to be lower for women and people with disabilities, whereas those who registered for larger and more expensive apartment units display higher levels of overall satisfaction. In column 2, we note that Orthodox Christians and Muslims are less satisfied than Protestants and other religious groups (the reference group). The same applies to individuals belonging to the Gurage ethnic group. Finally, there is a strong positive association between earnings prior to the lottery (in 2015) and overall life satisfaction.
are smaller

normal

with

A.G.

+/

percentages). “Overall life satisfaction,” “Financial distress index,” and “Psychological distress (K10)” are standardized using the mean and standard deviations of the losers group.

Table 4

<table>
<thead>
<tr>
<th>Overall life satisfaction</th>
<th>Financial distress</th>
<th>Psychological distress</th>
</tr>
</thead>
<tbody>
<tr>
<td>Winner</td>
<td>(1) 0.190***</td>
<td>(3) -0.116***</td>
</tr>
<tr>
<td></td>
<td>(0.036)</td>
<td>(0.034)</td>
</tr>
<tr>
<td>Controls</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>N</td>
<td>3049</td>
<td>3049</td>
</tr>
</tbody>
</table>

Note: The table details OLS estimates for the effect of winning the lottery on the main outcomes. Robust standard errors are in parentheses. P-values are ≤ 0.01***, ≤ 0.05**, and ≤ 0.1*. All regressions control for the strata fixed effects. The dependent variable is standardized using the mean and standard deviation of the loser group.

Columns 3 and 4 provide the correlates for financial distress and columns 5 and 6 provide those for psychological distress. These correlations are almost a mirror image of the results for life satisfaction, i.e., characteristics associated with lower levels of satisfaction are generally associated with higher levels of distress. However, the correlations appear weaker for psychological distress, which is not correlated with religion for instance. This suggests that the determinants of mental health to some extent may differ from those of life satisfaction and financial distress.

Our data point to a strong correlation between economic resources and well-being. But to what extent is this a causal relationship? Fig. 2 depicts the distribution of the main outcome variables for winners and losers. We can see that winners tend to report higher life satisfaction than losers, whereas the reverse is true for financial distress. By contrast, the distributions for psychological distress are more similar across winners and losers.

However, there are major differences between the two groups, some of which (the strata variables) are correlated with the probability of winning. To properly account for such differences, Table 4 provides treatment effect estimates for the main outcomes with and without controls. Fig. 3 presents the effects on the same outcomes and subindices graphically.21

We first report the effect of winning the lottery on the standardized satisfaction outcomes. As shown, overall life satisfaction increases with winning by 0.19 standard deviations (column 1). This effect is significant and is, for instance, larger than the gender gap in satisfaction. Considering the subindices, we identify similar effects on financial and neighborhood satisfaction, but with a smaller effect on satisfaction with leisure, whereas the effect on the domain “home” is almost twice as large as the effect on overall satisfaction.

We now provide the results for financial distress, where the overall effect on the additive index is 0.12 standard deviations (column 3). In considering the subindices, we see that winning the lottery affects all four outcomes related to financial distress in that winners

21 In Fig. 5, given the number of tests reported, we also adjust the confidence intervals for multiple hypothesis testing using the procedure of Benjamini and Hochberg (1995). We construct the adjusted confidence intervals based on adjusted critical values. With nine p-values, for example, the significance thresholds are 0.0055, 0.11, 0.017, ... , 0.05, instead of being 0.05 for all outcomes, and the corresponding critical values given by the normal distribution are 2.77, 2.54, 2.39, ... , 1.96, instead of 1.96 for all outcomes. We then define the adjusted 95% confidence interval as “estimate +/- (adjusted critical value) standard error.” We perform the corrections using “family of outcomes.” We have three distinct families in the figure: i.e., the life satisfaction in different domains (from health to work), the financial distress measures (from unforeseen situation to financial difficulties) and the 10 components of the Kessler psychological distress scale (from tired to worthless).
are less likely to have inadequate money for household expenses, to have delayed bills, and to have experienced financial difficulty. However, on one outcome the effect goes in the opposite direction. When asked about whether they would be able to raise a large amount of money (ETB 20,000) in a brief period of time if needed, seven percent more winners than losers reported that they would be unable to raise money. Although this may seem contradictory at first, it is consistent with the observation in Table 2 that winners have lower savings and more debt than losers. Indeed, most winners have already raised money by borrowing from friends and relatives, whereas losers are preparing to finance the down payment in case they win a lottery in the future.

Finally, in columns 5 and 6 and in the lower part of Fig. 3, we show the effect of winning on the K10 scale. The effect on the overall score is negative but small and not statistically significantly different from zero. The same applies to most of the individual items. We also note that the confidence intervals are relatively precise. Using an equivalence testing approach of two one-sided t-tests, and a five percent significance level, we can reject reductions in the K10 index equal to or greater than 0.1 standard deviations. Furthermore, when we compare the levels on the K10 scale for winners and losers while controlling for the strata variables, we see that losers score 15.46 on the full index in the range of 10–50 while the winners score 15.24. These results are shown in Appendix Table, where we
can also see that the coefficients for winning on mild, moderate, and severe mental distress are also negligible. We note that adding all controls does not change any of these findings, and in Appendix Section we show that this is also the case when adding optimal controls.

4.3. Robustness

Overall, we can see that winning the lottery resulted in large increases in wealth, reduced financial distress, and greater life satisfaction, but had no effect on psychological distress. These results are robust to different sets of control variables and alternative coding choices and the conclusions are similar if we conduct a bounds analysis accounting for selective nonresponse by winners (see Appendix Sections C, and, respectively).

We can also adjust our p-values for the fact that we are testing multiple hypotheses. We test two main hypotheses in this paper, but we also use these data to test for the effects on five different attitudes in a separate paper (Andersen et al., 2020).22 We prespecified an adjustment of the p-values for multiple testing using the false discovery rate method developed by Benjamini and Hochberg (1995). Despite the outcomes of the various analyses being quite different, we believe that it is prudent to adjust the p-values based on all tests with the same treatment and this is what we prespecified. With seven primary outcomes and a five percent significance level, our result with the lowest p-value should have a p-value lower than 0.007 (0.05/7). Our p-value for life satisfaction has a p-value lower than 0.001.

5. Discussion of mechanisms

In this section, we further test the plausible mechanisms through which winning the lottery affects well-being. Given the existing evidence on the strong correlations between housing conditions and measures of well-being (see e.g. Ludwig et al., 2012; Patel et al., 2006), we first investigate the extent to which access to better housing can explain the effects that we observe (Section 5.5.1). We find that access to better housing does not seem to play an important role in explaining the lottery’s effects because the effects are very similar for movers and nonmovers.

Second, we do a mediation analysis to estimate the importance of wealth, financial distress, and moving to a new house in explaining the effects of winning the lottery (Section 5.5.2). The analysis suggests that increases in wealth explain 62 percent of the total effects on life satisfaction, while reductions in financial distress explain 21 percent of the total effects. We find no mediation effect of moving.

Last, we explore the heterogeneity of the lottery’s effects along several dimensions, including initial levels of wealth (Section 5.5.3). The analysis of heterogeneous effects also confirms the importance of changes in wealth in explaining the lottery effects. In particular, we find that winning the lottery has a strong effect on life satisfaction among the “poor” participants but not among the “very poor” nor the “richer” participants.23

5.1. Movers and nonmovers

We mainly interpret the effects of winning the lottery in terms of a wealth effect. However, the observed effects on life satisfaction for the domains “home” and “neighborhood” suggest that moving to a better house may indeed drive part of the effect. To investigate this hypothesis further, Fig. 4 illustrates the treatment effect estimates for subgroups of winners: those who moved into their new apartment (“movers”) and those who did not (“nonmovers”).24 Because moving is not random, and Table reveals that movers are less likely to be born in Addis Ababa and more likely to have a partner, we have included a version of this figure in the Appendix, where the full set or a subset of optimally chosen control variables are included in the regressions.

We can see that the effect of winning on overall satisfaction is almost the same for movers and nonmovers. We also observe similar effects on neighborhood satisfaction, while movers exhibit higher satisfaction with their home, and nonmovers—who generally rent out their units—have higher financial satisfaction. This result suggests that even though winners have different priorities and spend their economic resources in dissimilar ways, the effects on overall life satisfaction are the same. The fact that both groups of winners have higher satisfaction with their homes and neighborhoods could be a compositional effect driven by the least satisfied people moving, and becoming happier with their housing conditions, which would lead to the nonmoving group also having higher satisfaction with their homes than the losers.

With respect to financial and psychological distress, the effects for movers and nonmovers are even more similar. While being aware of the risk of self-selection bias, we believe these findings strengthen the interpretation that the effects of winning the lottery are driven primarily by the changes in wealth rather than by moving to better houses and neighborhoods.

22 This paper is part of a larger project focusing on different effects of the Ethiopian housing lottery. We document the effects of the lottery on views about inequality and on support for redistribution in Andersen et al. (2020).
23 In Appendix Section, we also show the results of instrumental variable analyses where we use winning the lottery as an instrument for wealth or financial distress. The instrumental variable approach provides consistent evidence of substantial effects of wealth and financial distress. As the exclusion restriction is unlikely to hold in these analyses, i.e., as winning is likely to affect well-being in ways other than through wealth or financial distress, we emphasize that these results should be interpreted cautiously.
24 “Movers” are all winners who actually moved into the apartment they won and “nonmovers” are all other winners, including those who chose to rent out the apartment they won.
5.2. Mediation analysis

To more directly test whether wealth and/or financial distress, as well as moving, are mediating the relationship between winning and well-being we conduct non preregistered mediation analyses. These analyses consist of estimating three sets of regressions: 1) the reduced form effect of winning on well-being; 2) the reduced form effect of winning on the mediator; and 3) the effect of winning on well-being while controlling for the mediator. Under the assumption that the error terms in 2 and 3 are uncorrelated (the sequential ignorability assumption) one can decompose the total effect in 1) into the direct effects of winning and the average causal mediation effect (ACME) (Imai et al., 2011). However, the sequential ignorability assumption is strong and not directly testable. The assumption implies that there should be no unobservable variable that affects both well-being and the mediators. For example, the assumption is violated in the analysis of wealth as a mediator if family background (which is not included in our model) affects both wealth and well-being. Hence, the results from the mediation analysis should be seen as suggestive and more descriptive than the reduced form results.

Fig. 4. Effect on the main outcomes and subindices by mover status. Note: The figure depicts the estimated effects of winning on the main outcomes and the subindices for movers and nonmovers, respectively. The bars denote 95 percent confidence intervals around the point estimates (the shorter thicker intervals are not adjusted for multiple hypothesis testing and the longer thinner intervals are adjusted using the Benjamini–Hochberg procedure). All estimations include the strata variables but include no additional controls.
We start with investigating how the effects of the lottery on satisfaction are mediated by wealth. In the sample with nonmissing wealth, the reduced form effect of winning is equal to 0.14. The average effect of winning the lottery on the outcome that operates through wealth is 0.09 (ACME). The estimate of the direct effect of winning the lottery is 0.05. These results are displayed in Table 5, panel A. The direct effect is not statistically significant and the share of the total effect that is mediated by wealth is high, at 62 percent. When we add the vector of baseline controls (Appendix Table ), the direct effect estimate is even lower and the share of the total effect that is mediated by wealth rises to 73 percent.

Using financial distress as a mediator yields an ACME of 0.04 (Table, panel B). The direct effect is now statistically significant and the share of the effect that is mediated by financial distress is 21 percent. This share falls to 18 percent when we add the vector of control variables (Appendix Table ).

In panel C, we use moving as a mediator. There is no statistically significant mediation effect. Hence, moving does not appear to be a mediator for the effect of winning on satisfaction. We do not report the tables for psychological distress as there is no average effect of winning that can be meaningfully decomposed.

5.3. Heterogeneous effects

Our findings have shown that winning the lottery improves overall life satisfaction, but that the effects on psychological well-being are less clear. A natural question to ask is whether these overall effects are driven (or concealed) by differential effects for different groups of winners. In this section, we test for heterogeneous treatment effects across pre-lottery earnings and the strata variables.25 First, we split the sample by the median of earnings in 2015 (before the lottery). We create a dummy variable equal to 1 for the richest 50 percent and zero otherwise, and interact this dummy with winning. As seen in Table 6, richer people have significantly higher life satisfaction and lower psychological distress on average (columns 1 and 4). Winning the lottery increases overall life satisfaction by 0.218 standard deviations among below-median earners (the reference group). This roughly corresponds to the difference in satisfaction between above- and below-median earners (0.233). The sign of the interaction terms suggests the effect of winning is smaller among above-median earners, but it is not statistically significant. Indeed, we can reject that the effect is zero even for this group (p = 0.049). We observe no significant effects on psychological distress for either group.

To explore heterogeneity across more detailed earning categories we split the sample by earnings quartiles. We let the richest 25 percent be the baseline category and we interact the other categories with winning. We see several interesting patterns in columns 2 and 5. The effect of winning on satisfaction for the richest people is small (0.039) and not statistically significant. The quartile dummies indicate that the richest were more satisfied and less distressed than the lower quartiles, leaving less scope for improvements in this group. Interestingly, we find that winning increases satisfaction and lowers distress among the poor (quartile 2), but not the poorest (quartile 1). A plausible interpretation is that the poorest winners face additional liquidity constraints due to the down payment.

We also explore heterogeneity by interacting winning with our different strata variables (columns 3 and 6). Interestingly, we find that winning the lottery has particularly beneficial effects among disabled people, who otherwise exhibit much lower life satisfaction.

---

25 We also further explored heterogeneity using the “generic machine learning approach” in Chernozhukov et al. (2018). This method includes an omnibus test of heterogeneity in the treatment effects, and we cannot reject the null hypothesis that there is no heterogeneity overall.
Table 6
Heterogeneous treatment effects.

<table>
<thead>
<tr>
<th></th>
<th>Overall life satisfaction</th>
<th>Psychological distress</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Winner</td>
<td>0.218***</td>
<td>0.039</td>
</tr>
<tr>
<td></td>
<td>(0.053)</td>
<td>(0.067)</td>
</tr>
<tr>
<td>Winner × Above median</td>
<td>-0.113</td>
<td>0.070</td>
</tr>
<tr>
<td>Above median</td>
<td>0.233***</td>
<td>-0.162***</td>
</tr>
<tr>
<td>Winner × Quartile 1</td>
<td>0.017</td>
<td>-0.019</td>
</tr>
<tr>
<td>Winner × Quartile 2</td>
<td>0.318**</td>
<td>-0.191</td>
</tr>
<tr>
<td>Winner × Quartile 3</td>
<td>0.145</td>
<td>-0.104</td>
</tr>
<tr>
<td>Quartile 1</td>
<td>-0.367</td>
<td>0.303</td>
</tr>
<tr>
<td>Quartile 2</td>
<td>-0.337***</td>
<td>0.157**</td>
</tr>
<tr>
<td>Quartile 3</td>
<td>-0.235***</td>
<td>0.120*</td>
</tr>
<tr>
<td>Winner × Female</td>
<td>0.057</td>
<td>-0.047</td>
</tr>
<tr>
<td>Winner × Public</td>
<td>0.034</td>
<td>-0.061</td>
</tr>
<tr>
<td>Winner × Disabled</td>
<td>1.226***</td>
<td>-0.906**</td>
</tr>
<tr>
<td>Winner × One bedroom</td>
<td>0.124</td>
<td>0.082</td>
</tr>
<tr>
<td>Winner × Two bedroom</td>
<td>0.029</td>
<td>0.114</td>
</tr>
<tr>
<td>Female</td>
<td>-0.113***</td>
<td>-0.083*</td>
</tr>
<tr>
<td></td>
<td>(0.043)</td>
<td>(0.043)</td>
</tr>
<tr>
<td>Public</td>
<td>0.054</td>
<td>0.034</td>
</tr>
<tr>
<td></td>
<td>(0.043)</td>
<td>(0.043)</td>
</tr>
<tr>
<td>Disabled</td>
<td>-0.312***</td>
<td>-0.279***</td>
</tr>
<tr>
<td></td>
<td>(0.099)</td>
<td>(0.098)</td>
</tr>
<tr>
<td>One bedroom</td>
<td>0.225***</td>
<td>0.227***</td>
</tr>
<tr>
<td></td>
<td>(0.048)</td>
<td>(0.048)</td>
</tr>
<tr>
<td>Two bedroom</td>
<td>0.263***</td>
<td>0.275***</td>
</tr>
<tr>
<td></td>
<td>(0.056)</td>
<td>(0.056)</td>
</tr>
<tr>
<td>Strata</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Controls</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>N</td>
<td>3049</td>
<td>3049</td>
</tr>
</tbody>
</table>

Note: The table details OLS estimates for the effect of winning the lottery for different groups. Robust standard errors are in parentheses. P-values are ≤ 0.01***, ≤ 0.05**, and ≤ 0.1*. All regressions control for the strata fixed effects. The dependent variable is standardized using the mean and standard deviation of the loser group.

(by 1.47 standard deviations) and higher levels of psychological distress (by 1.29 standard deviations) than the general population. In this group, winning increases life satisfaction by 1.29 standard deviations and reduces psychological distress by 0.92 standard deviations (p = 0.038). We do not find any significant heterogeneity for the other strata variables. 26

6. Conclusion

The question of the extent to which material conditions affect well-being has a long history in the social sciences. On the one hand, economic resources can be used to obtain desired goods and services. It would seem obvious that they should increase well-being. On the other hand, not everything of value in life can be bought and humans have a remarkable ability to adapt to their material circumstances. Adding to the lack of clear theoretical predictions is that it is not straightforward whether material conditions affect well-being, whether well-being affects material conditions, or whether there is some third factor affecting both.

26 In the pre-analysis plan, we also stated that we would use the variation in housing prices across areas to assess whether our results depend on the size of the wealth shock. However, this analysis has proven infeasible because 95 percent of the winners were assigned to only two areas, for which the estimated housing prices are almost identical. With such limited variation, using the dispersion across areas is not possible.
Moreover, there is currently much attention drawn to the prevalence of CMDs in low- and middle-income countries, which was ignored for a long time by the scientific community. There is growing, but still limited, evidence on the effects that poverty has on mental health in such contexts.

We identify the causal effects of winning a housing lottery in Ethiopia on different well-being dimensions. We first report the important effects that winning has on wealth. We then show that winning increases life satisfaction but does not affect psychological distress on average. Our estimates are very precise, and we can reject very small effects (0.1 standard deviations) on psychological distress. When we investigate subindices of life satisfaction, we see that the effects are driven by greater satisfaction with housing, neighborhood, and personal finances. We also find marked reductions in financial distress.

Winning the lottery has effects other than increasing wealth; it particularly also affects housing conditions and neighborhood characteristics. Given that only 30 percent of the lottery winners actually moved into their new property, and that we find similar results for both movers and nonmovers, we believe that the effects we identify on life satisfaction are due to a wealth effect. A mediation analysis indicates that most of the effect occurs via wealth, less of the effect occurs via financial distress. On the other hand, we cannot reject that moving does not explain the effects of winning.

Earlier studies that identified the effects of changes in material conditions mainly focus on rich countries (e.g., Apouey and Clark, 2015; Gardner and Oswald, 2007; Lachowska, 2017; Lindqvist et al., 2020; Schandt, 2018) and generally find that material resources increase happiness and life satisfaction. In Sweden, Lindqvist et al. (2020) also find that winning a lottery does not improve mental health. The absence of effects on mental health in Sweden could be explained by the country’s comprehensive welfare system, which guarantees economic security and high-quality health care to its citizens (Ridley et al., 2020). From this perspective, it is remarkable that we find qualitatively the same results in a context where there is no welfare state or economic security. Our evidence may hint at the presence of different factors determining mental health and life satisfaction, as also suggested in the literature (Kahneman and Deaton, 2010; Weich et al., 2011). In particular, life satisfaction has lower heritability than other aspects of well-being, it is more influenced by environmental factors than many other well-being dimensions (Barrels, 2015; Rayssamb and Nes, 2018; Rayssamb et al., 2018). The differences in these results may also be explained by winning “only” affecting material aspects of life satisfaction, as shown in analyses of subindices, while not affecting more intangible aspects of life satisfaction and psychological distress.

However, other studies from low-income countries have identified the positive effects of economic resources on mental health and well-being. These studies typically investigate the effects of cash transfers or antipoverty programs on the extremely poor (McGuire et al., 2020; Ridley et al., 2020; Zimmerman et al., 2021, for recent reviews, see). One important difference in our case is that the previously studied interventions are targeting very poor individuals. We also note that McGuire et al. (2020) document generally smaller effects on mental health than on life satisfaction. While our respondents are certainly not rich, they are neither among the poorest Ethiopians. In addition, we find that poor individuals, but not the poorest, experience the most positive effects of winning in our setting. This may hint at specific factors of our lottery being important for the results, such as winners having increased debt and more difficulty in borrowing (more) money in the case of an emergency. It is possible that the results on psychological distress would be larger if winning the lottery did not entail any down payment.

Policymakers may find our results useful. The fact that the benefits from the housing lottery seem to come from increases in (real estate) wealth, rather than from moving to a new house, is particularly relevant for social housing programs and formalization of property rights for example, two policies that have become increasingly important in low-income countries. Importantly, the lack of a general reduction in psychological distress, despite that large increase in wealth, in a population with a high prevalence of mild to severe mental disorders is worrying and calls for testing alternative policies to improve mental health in this context. We trust that future studies continue to investigate the effects of material conditions on mental health and well-being for different types of populations and different types of interventions so that we can reach a better understanding of this important relationship.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jhealeco.2022.102619.

CRediT authorship contribution statement

Asbjørn G. Andersen: Conceptualization, Methodology, Formal analysis, Writing – original draft. Andreas Kotsadam: Conceptualization, Methodology, Formal analysis, Writing – original draft, Writing – review & editing, Funding acquisition. Vincent Somville: Conceptualization, Methodology, Formal analysis, Writing – original draft, Writing – review & editing, Funding acquisition.

References


Cerhan, Z., Costello, Konstantopoulou, a doi: 10.1111/j.1467-9442.2006.00459.x


Chapter 5:

Does wealth reduce support for redistribution? 
Evidence from an Ethiopian housing lottery