THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par l'Université Toulouse 1 Capitole

Présentée et soutenue par

Jian Xin Hansel TEO

Le 5 septembre 2022

Essays on the economics of insurance markets

Ecole doctorale : **TSE - Toulouse Sciences Economiques**

Spécialité : **Sciences Economiques - Toulouse**

Unité de recherche :
TSE-R - Toulouse School of Economics - Recherche

Thèse dirigée par
Helmuth CREMER et Jean-Marie Lozachmeur

Jury

M. Mathieu LEFEBVRE, Rapporteur
M. Pierre PESTIEAU, Rapporteur
Mme Chiara CANTA, Examinatrice
M. Emmanuel THIBAULT, Examinateur
M. Helmuth CREMER, Directeur de thèse
M. Jean-Marie LOZACHMEUR, Co-directeur de thèse

Essays on the Economics of Insurance Markets

Hansel Teo

Toulouse School of Economics
Universite Toulouse Capitole 1
Toulouse, France
June 2022

Abstract

What affects people's decision to purchase insurance and what determines if they are insured? This thesis collects three chapters which explore these questions from different perspectives. In Chapter One, I adopt the sufficient-statistics approach to quantify the consumption-smoothing value of long-term care insurance. My analysis extends the current framework to allow for adjustment costs in consumption and derives a generalised implementation formula for measuring the marginal value of insurance. I then implement the derived measure by using panel data from the Health and Retirement Survey to estimate the implied value of insurance against nursing home episodes. On average, I find that households' food consumption drops sharply by about 26 per cent during a nursing home episode, while the probability of home liquidation increases by nine percentage points. These estimates imply uninsured people are willing to pay a 60 per cent premium over actuarially fair rates for insurance against nursing home expenses. In Chapter Two, I propose and analyse a monopoly insurance contracting model with unobserved heterogeneity in both risk and preferences for formal insurance. I show that monopoly optima can have partial take-up and adverse selection along either, or both, the intensive and extensive margins. The particular configuration at an optimum depends on the degree of heterogeneity in participation preferences. With a low degree of heterogeneity, all individuals are insured and there is only selection between contracts. When there is a high degree of heterogeneity, there is partial take-up and adverse selection occurs simultaneously between contracts and in take-up. In Chapter Three, I study how imperfect competition affects equilibrium insurance allocations by analysing a model of horizontally-differentiated insurers. I show that equilibrium in this model always features risk separation, with the equilibrium allocation characterised by conditions which capture intensive and extensive margin trade-offs in profit. Consumer surplus is strictly increasing and while the level of coverage provided to low-risks is strictly decreasing in the degree of competition.

Acknowledgements

I would like to thank my PhD supervisors, Jean-Marie Lozachmeur and Helmuth Cremer for their support throughout my time at Toulouse. Special thanks also to Nicolas Werquin for his time and encouragement, particularly when the going got tough. My time at Toulouse would not have been the same without those easily overlooked discussions and interactions with colleagues and I would like to express my appreciation for all who have had to "put up" with me.

Finally, and most importantly, I would like to thank my parents, who brought me to this world, and the aunt and uncle, who raised me and supported me all this time. I would certainly have not made it this far without them and am immeasurably grateful for their having them in my life. This work is dedicated to you.

Contents

Chapter 1

The value of long-term care insurance: A sufficient-statistics approach

1.1 Introduction

Long-term care dependency risk - the risk of encountering a prolonged episode of poor health and requiring assistance with daily living - is increasingly important given the ageing populations in many developed countries. However, unlike acute medical care, long-term care expenses are not currently covered by social insurance programmes in most countries. Furthermore, the take-up of private long-term care insurance (LTCI) is very low. The question of who should cover long term care expenses and how best to do so are hence pressing policy issues. Is asymmetric information responsible for the low rate of private LTCI? Or do people simply not value having formal LTCI? If the government were to step in, would the benefit of providing universal public LTCI outweigh its associated costs? A key input required to answer any of these questions is the value of LTCI.

This paper seeks to quantify the marginal value of LTCI. Conceptually, this value is the marginal rate of substitution for wealth between good health and dependency. More concretely, this can also be interpreted as the increase in premiums an individual would be willing to pay for a marginal increase in indemnities.

To measure this marginal value of LTCI, we adopt an approach developed in the social insurance literature which infers the marginal value of insurance against a shock by observing the consumption responses to these shocks. This "consumption-based" approach is part of a broader "sufficient-statistics" literature that quantifies the welfare effect of policy changes by estimating high-level elasticities rather than structural primitives.

In applying the approach to our context, we extend the existing consumption-based framework to account for goods that incur adjustment costs. Such "consumption commitments" are particularly relevant to the context of LTCI, where housing choices are important. However, the presence of this consumption category poses a well-known challenge to the existing consumption-based implementation framework. Intuitively, the consumption-based approach takes an estimate of consumption response to a shock and weights it by a measure of risk aversion to obtain an implied marginal value of insurance against this shock. However, when there are both fully flexible and committed consumption goods, the standard framework is unclear on which of these consumption categories is relevant and what the corresponding risk aversion measure should be. [1]

To address the methodological issue of consumption commitments, the first part of this paper sets up a two-good model with flexible ("food") and committed ("housing") consumption goods, and uses it to derive a generalised consumption-based implementation formula. In the second part of the paper, we implement our generalised measure on panel data from the Health and Retirement Study (HRS) to obtain an estimate for the marginal value of insurance against nursing home episodes.

Our theoretical analysis yields two critical insights. First, we need to be able to distinguish between flexible and committed consumption responses in the data. This is because the logic of inferring the value of insurance from marginal changes in consumption is valid only for the flexible consumption good. Intuitively, responses in housing, the good with adjustment costs, are "contaminated" because they are simultaneously influenced by both the shock itself and also by adjustment costs. Thus, estimated responses in housing consumption do not cleanly reflect the effect of the shock.

The second key insight is that the presence of adjustment costs induces a systematic difference in the benefit of insurance between people who do make costly consumption adjustments ("adjusters")versus those who do not ("non-adjusters"). Intuitively, by bearing the adjustment cost, adjusters can absorb a negative income shock by reducing both food and housing consumption. On the other hand, non-adjusters have to absorb the full impact of any shock through food consumption only.

[1] For example, if consumption responses are measured using a basket of both flexible and committed consumption goods then the curvature of utility depends on whether or not committed consumption is adjusted (see Chetty and Szeidl (2007)).

Our generalised measure of the value of insurance accounts for these implications by using flexible consumption responses of adjusters and non-adjusters and adjusting and taking a weighted sum of consumption responses of these two groups to account for the heterogeneity between them. At a practical level, our generalised representation highlights that flexible consumption is the essential consumption category for implementation and that the corresponding curvature measure is the one that assumes fully adjustable consumption.

To understand the implications of ignoring consumption commitments when they are relevant, we use our model to study the errors which arise from violating the two key implications discussed above. This exercise also allows us to relate our present findings to the existing literature using the consumption-based approach.

We show that when food consumption is taken as the focal consumption good, a naive implementation that ignores heterogeneity in insurance value between adjusters and non-adjusters results in underestimation that is of second-order magnitude. In contrast, errors of first-order magnitude can arise when researchers observe only the sum of food and housing consumption. Intuitively, housing consumption changes among adjusters inflate the observed consumption response. In turn, this leads to an upward bias in the implied marginal value of insurance if the same curvature parameter is used to scale the average consumption across the population. Put another way, ignoring the need to focus on only flexible consumption responses results in "contamination" of consumption response measure for adjusters and failure to distinguish between adjusters and non-adjusters means that this contamination cannot be accounted for.

Based on our theoretical framework, empirical implementation of our measure of the value of insurance can be summarised as seeking the answers to the following three questions. First, do people's flexible consumption respond to dependency shocks and, if so, by how much? Second, do people respond to dependency shocks by making the costly adjustment of moving homes and, if so, how much? Third, how do the changes in flexible consumption differ between people who also move in response to a dependency shock versus those who do not?

To answer these questions, our empirical application uses the panel structure of the HRS to track the dynamics of household food consumption expenditure and housing market transactions in the six-year window around an elderly household member's nursing home entry.

We begin by first ignoring heterogeneity between adjusters and non-adjusters. This amounts to applying the standard consumption-based implementation directly to our context. Using household food expenditure as our measure of flexible consumption, we document sharp consumption decreases at the onset of a nursing home spell and persistently reduced consumption levels throughout the length of the spell. To translate these changes in reported consumption at each survey wave to a measure of flow consumption response, we adopt the simple parametric specification used in Landais and Spinnewijn (2019). This procedure implies a flow consumption drop of about 26.7 percent in response to a nursing home episode. In turn, this implies that households are willing to pay at least a 34 percent markup over the actuarially fair rate for the first unit of insurance against a nursing home episode.

We find that nursing home episodes are associated with large and immediate increases in out-of-pocket medical expenses equating to about 6.5 times the amount of pre-event spending on food consumption. These medical expenses are composed predominantly of nursing home care expenses and persist throughout a nursing home spell. In contrast, household income does not respond significantly to the onset of a nursing home episode. Also, to the extent that any compensating increase in income is observed, these increases occur at least two years after the entry into nursing home care. These patterns highlight the large mismatch between care-related expenses and compensating income transfers and reiterate the fact that nursing home episodes are associated with significant negative income shocks.

To study households' extensive margin response in housing consumption, we focus on observations of home sales. These make up the majority of housing transactions in our sample. We find that while home sales are infrequent amongst our sample of elderly, a nursing home entry increases households' probability of a selling their home by around 9.0 percentage points. This sharp increase occurs at the immediate onset of a nursing home spell and remains positive in the subsequent survey wave before returning to zero.

Having shown that households who encounter a nursing home episode adjust both their food and housing consumption, we turn to study heterogeneity in food consumption responses between households who make extensive margin housing adjustments and those who do not. To do so, we group households into those who sell their homes in the immediate nursing home entry wave ("movers') and those who do not ("non-movers"). Then, we repeat our analysis while allowing for differences in consumption responses between these two groups. This exercise finds that while non-movers experience a 43.8 percent flow consumption decrease due to a nursing home episode, movers face a smaller 12.4 percent flow consumption drop. This difference in consumption responses maps to a corresponding gap in the implied marginal value of insurance. Under the same risk aversion assumptions as before, "non-movers" are willing to pay a 63 percent markup over the actuarially fair rate for insurance. In contrast, "movers" are only willing to pay a 14 percent markup.

To obtain an estimate of the aggregate marginal value of insurance, we use our weighted sum formula along with these estimates. This yields an estimate of 1.59, which implies that there are significant consumption-smoothing benefits from insurance against nursing home episodes. Unsurprisingly, this implied value of insurance is increasing in risk aversion. However, our analysis highlights an important nuance: much of the consumption-smoothing gains from insurance accrue to non-movers. This finding suggests potential efficiency gains from conditioning LTCI on observable measures of households' capacity to adjust consumption.

1.1.1 Related literature

This paper speaks to two distinct areas of work. First is the increasingly rich body of work analysing the demand for and provision of private LTCI. Starting from Brown and Finkelstein (2008), the literature studying the demand for private LTCI has largely considered a life-cycle model and adopted a calibration or structural estimation approach. From this literature, a number of papers have highlighted various factors which could explain why individuals may not want to fully insure against dependency risk using LTCI. These include interaction with public LTCI insurance (Brown and Finkelstein, 2008); substitution within a financial portfolio of various insurance instruments (Koijen et al., 2016); substitution between LTCI and housing equity Davidoff (2010); the interaction between bequest motives and dependency risk(Lockwood, 2018); the discrepancy between real-world LTCI products and "abstract" LTCI studied in models (Ameriks et al., 2016) and, more recently, knowledge about LTCI products and dependency risk (Boyer et al., 2020). With respect to this literature, the present paper contributes to our knowledge about the demand for LTCI by proposing a measure for the value of LTCI which is consistent with the various characteristics of the economic environment highlighted and whose conclusion is robust to different assumptions about underlying risk processes.

In addition to the demand-side explanations above, the thinness of real-world LTCI markets can also be attributed to supply-side factors. In this regard, Finkelstein and McGarry (2006) find that private information about both one's risk and risk preferences affect LTCI can lead to selection in take-up. Building on this, Hendren (2013) also finds that the cost that asymmetric information imposes on insurers is large enough to explain why some individuals are rejected and why certain segments of the LTCI market are completely non-existent. In addition, recent work has studied how asymmetric information can interact with means-tested public insurance to depress equilibrium private LTCI take-up and coverage (Braun et al., 2019). To the extent that asymmetric information imposes poses a challenge to private market LTCI provision, an important policy question is whether there is a role for public provision. In turn, an essential input to such an evaluation is a measure of the benefit (or value) of providing insurance. In this regard, our contribution is to provide a transparent and tractable measure of the value of LTCI.

Outside of LTCI, this paper also builds on the literature using the "sufficient-statistics" approach to study the value of insurance. Beginning from Gruber (1997) and building on the theoretical insight from Chetty (2006), a series of papers have derived implied values for insurance against unemployment (Hendren, 2017; Kolsrud et al., 2018; Ganong and Noel, 2019; Landais and Spinnewijn, 2019) and disability (Meyer and Mok, 2019) by using the "consumption-based approach" which maps from estimated consumption responses to uninsured shocks to a marginal value of insurance against those shocks. However, as mentioned above, a drawback of this approach is its sensitivity to the assumed curvature of preferences which, in turn, is influenced by the presence of committed consumption (Chetty and Szeidl, 2007). In this respect, the present paper contributes to this literature by extending the theoretical framework to account explicitly for consumption commitments and showing how this implies the need to observe additional consumption responses beyond that required in the standard model.

Finally, this paper adds to the recent literature which analyses households' dynamic response to health shocks using an event-study design (Dobkin et al., 2018; Meyer and Mok, 2019; Fadlon and Nielsen, 2019; Kolsrud et al., 2019). In this regard, the present paper is the first to analyse the effect of nursing home episodes on households' consumption, medical expenditures and income.

The remainder of this paper proceeds as follows: in the next section, we set up and analyse a theoretical model with two consumption goods (one flexible one committed) and dependency risk. Next, Section 1.3 describes the data and empirical strategy used to implement our framework. Sections 1.4 and 1.5 discuss the estimated responses in flexible and housing consumption, respectively. Section 1.6 consolidates our earlier findings by examining the heterogeneity in consumption response on both sides of the housing extensive margin and deriving implied values of LTCI using our sufficient-statistics measure. Section 1.8 concludes.

1.2 Conceptual model

In this section, we set up a model with flexible and committed consumption goods. Within this setup, we derive a representation of the marginal value of insurance against becoming dependent as a function of consumption responses to these shocks. Finally, we use this model to discuss the implications of ignoring committed consumption when they are relevant.

1.2.1 Setup

Preferences and risk Consider a finite, discrete-time setting in which agents can be in one of three health states, denoted h_t, in each period: good health (G), dependent (B) or dead (D). In each period, agents consume two goods, food (c) and housing (x), and have instantaneous preferences given by $u_h(c,x)$. We assume that u is increasing and concave in (c,x) and that $\frac{\partial^2 u_h}{\partial c \partial x} \geq 0$.[2] The latter assumption implies that we allow for preferences over (c,x) to be complementary.[3] Lastly, subscripts, $h \in G,B,D$ indicate preferences can depend on health-states.

Agent's problem Focus on a single agent. The agent arrives at the start of each period, t, with wealth, w_t, and a prior level of housing consumption, x_{t-1}, before subsequently observing h_t.

Health realisations have implications both on resources and preferences. Specifically, dependency incurs expenses and also possibly affects preferences over consumption. We model expenses as an exogenous negative income shock, L, and assume that it is constant. The assumption of constant per-period expenses is consistent with real-world long-term care expenses, but our model can nevertheless be extended to analyse time-varying in-

[2]This is satisfied by CES utility over (c,x) and separable power utility with CRRA ≥ 1.

[3]Where complementarity here is taken to mean that marginal utility in c is increasing in x and vice-versa.

come shocks.[4] However, this assumption abstracts from the choice of the type of care used, such as the choice between formal and informal care. It also abstracts from choice of "quality". Our analysis can thus be interpreted as the value of insurance conditional on choice over the type of care.

Let q_t denote the period-t price of housing consumption (relative to food), r_t the net return on wealth and β the agent's discount factor. Given prices, current wealth, w_t, prior housing consumption, x_{t-1}, and health history, $(h_1,\ldots,h_t)$, an agent chooses current food and housing consumption (c_t,x_t) and saves any unconsumed resources.

Assume that a choice of $x_t \neq x_{t-1}$ incurs a known utility cost $k \geq 0$.[5] Housing consumption is "committed" because an adjustment from its previous level incurs a cost. To fix terminology, we refer to a choice of $x_t \neq x_{t-1}$ as "moving" and $x_t = x_{t-1}$ as "staying put".[6] Formally, given the vector of state variables, $\omega_t = (w_t, x_{t-1}, h_t, \ldots, h_1)$, the agent's problem at period-t reads

$$v(\omega_t) = \max_{c,x} \quad u_h(c,x) - k \times \mathbf{1}(x \neq x_{t-1}) + \beta E_t[v(\omega_{t+1})] \tag{1.1}$$
$$s.t. \quad \begin{cases} c + q_t x + \frac{w_{t+1}}{r_t} \leq w_t + q_t x_{t-1} & \text{if } h = G \\ c + q_t x + \frac{w_{t+1}}{r_t} \leq w_t - L + q_t x_{t-1} & \text{if } h = B \end{cases}$$

Let $\pi_{B,1}$ and $\pi_{G,1}$ denote the unconditional probabilities of becoming dependent and of remaining good health at period-t respectively. Substituting recursively, the agent's ex-

[4] *Within* the same class of care (e.g. residential nursing, non-residential daytime or in-home care) daily/monthly care costs do not vary greatly and individuals' total expenses are driven largely by the duration of dependency. Thus, unlike acute medical care, "severity" in the context of LTC is more typically interpreted as the duration of time in care

[5] For our baseline setting, we assume adjustment costs are constant for each agent. However, in general, we can allow adjustment costs at t to be a function of ω_t, allowing an agent's cost of moving to depend on their current wealth, prior housing, and their current and past health status. In this context, we assume that the mapping from ω_t to the level of adjustment cost is known to the agent. That is, contingent costs of adjustment are known ex-ante to the agent.

[6] This terminology can be interpreted as decisions made by renters and homeowners as follows: 1. Renters arrive at period-t with wealth, w_t, and existing accommodation which provides x_{t-1}-units of housing consumption (e.g. 25 square-metres). A decision to adjust, $x_t \neq x_{t-1}$, incurs hassle costs, k, which we denominate in utility. 2. Homeowners, arrive at period-t with "liquid" wealth, w_t, and a home which provides x_{t-1}-units of housing consumption. $q_t x_{t-1}$ can be interpreted as the rental-equivalent of current housing. Our formulation implicitly implies that financial markets are complete and households are optimising with respect to home-equity release instruments. That is, the value of home equity is considered to be fully liquid and contained within w_t.

ante expected utility is

$$V = \pi_{B,1} v(w_1, x_0, B) + \pi_{G,1} v(w_1, x_0, G) + \pi_{D,1} v(w_1, x_0, D) \tag{1.2}$$

where $\pi_{D,1} = 1 - \pi_{B,1} - \pi_{G,1}$. We assume that the utility when dead, $v(w_1, x_0, D)$, is a non-decreasing function of w_t, x_0. This nests the common cases with no bequest preferences and bequests as a luxury good.

To pin down (w_1, x_0), we assume that (w_1, x_0) are chosen ex-ante, at time $t = 0$, to maximise V subject to the constraint, $w_1 + q_0 x_0 = w_0$, where $w_0 > 0$ is a given endowment.

Definition of LTCI

We define LTCI to be a contract which specifies an indemnity, $I_t \geq 0$, to be paid by the insurer whenever $h_t = B$ and which requires premium, $P_t \geq 0$, to be paid by the agent whenever $h_t = G$. The contract is purchased at $t = 0$ and enforced with full commitment until death. We focus on contracts with $I_t = I$ and $P_t = P$ for all $t = 1, \ldots, T$. That is, we restrict contracts to have constant benefits over an unlimited duration.

1.2.2 The value of intra-temporal consumption-smoothing

Within the setting described above, we study the value of insurance against dependency. To focus on the implications of introducing consumption commitments, the remainder of our analysis studies the static case. This obtains as a special case of the above setup if we assume that h_1 is absorbing and $\beta = 1$. The first assumption implies that all uncertainty is resolved at $t = 1$ and the second assumption implies that the optimal consumption path is flat for $t = 1, \ldots, T$. These, in turn, imply that any choice to move takes place at $t = 1$.[7] In light of our assumptions, we hereafter drop all time subscripts to ease notation.

Given any (I, P), consider a marginal increase in I and let WTP be the corresponding increase in P that leaves an agent indifferent. That is, WTP is the agent's marginal willingness to pay for LTCI. Let $v_h \equiv v(w_1, x_0, h)$ denote the agent's utility in health-h evaluated at their optimal consumption bundle. Formally,

$$WTP = -\frac{dV}{dI} \Big/ \frac{dV}{dP} = \frac{\pi_B}{\pi_G} \frac{\frac{\partial v_B}{\partial w}}{\frac{\partial v_G}{\partial w}}$$

[7] The extension to a dynamic setting is relegated to the supplemental appendix. Please contact the author for a copy.

The second equality obtains from applying the envelope condition. This equation says that WTP consists of two components: the utility value of transferring a marginal unit of wealth from the state in which the agent is healthy to the state in which they are dependent, and the ex-ante probabilities of each state occurring.

For the remainder of our analysis, we focus on the utility value of smoothing consumption across health-states. To do so, we define an agent's marginal rate of substitution of wealth between good health and dependency as $MRS = \frac{\partial v_B}{\partial w} / \frac{\partial v_G}{\partial w}$ and use this as our measure of the marginal value of insurance. Focussing on MRS allows us to make statements about the marginal value of insurance that does not depend directly on the underlying structure of risk.[8] To relate MRS to WTP, we assume that agents' risk perceptions coincide with their actuarial risk. Then, $MRS > (<) 1$ implies that agents are willing to pay more (less) than the actuarially-fair rate for a marginal increase in coverage.

Implications of consumption commitments for individuals

Consumption commitments affect an agent's MRS due to two fundamental reasons. First, because of adjustment costs, an the agent optimally moves if and only if the benefit from freely reallocating resources between food and housing exceeds the cost of adjustment. This means there is an extensive margin in housing adjustment since the agent simultaneously decides *whether* to move and, if so, what *level* of housing consumption to choose. Second, an agent's housing choice affects the marginal value of wealth through the budget constraint. In particular, while an agent who chooses $x \neq x_0$ ("an adjuster") has $w + qx_0$ of resources to freely allocate between food and housing, an agent who chooses $x = x_0$ ("a non-adjuster") can only allocate w to food consumption.

Together, these imply that the value of transferring resources between health-states differs depending on whether the agent moves or stays put. Since agents know their the adjustment cost ex-ante, they also know their health-contingent housing choice. Hence, an agent's marginal value of insurance, MRS, depends on their ex-post housing adjustment.[9]

[8] This allows us to address that fact that existing estimates of the willingness to pay for insurance are sensitive the underlying risk model. See Friedberg et al. (2014).

[9] In general, there are 2^2 possible contingent move choices: (move if $h = B$, move if $h = G$), (move if $h = B$, stay put if $h = G$), (stay put if $h = B$, move if $h = G$) and (stay put if $h = B$, stay put if $h = G$).

Inferring the value of insurance when adjustment costs are heterogeneous

An individual agent's marginal value of insurance depends on their housing consumption response to dependency. However, because insurance contracts cannot be conditioned on agents' ex-post consumption choices, the policy-relevant measure of MRS averages over adjusters and non-adjusters.[10] We now consider how to infer this value when adjustment costs are unobservable to the researcher.[11]

Assume k's are unobservable but are known to be independent and identically distributed in the population according to the cdf $G(k)$. As a shorthand, we refer to variables which integrate over the distribution of k as a "population" variable and use calligraphic script to denote such objects. $\mathcal{V} = \int V(k)dG(k)$, the "population" ex-ante expected utility, is the relevant objective for a utilitarian social welfare function or a private insurer facing a population with unobservable adjustment costs.[12]

We are interested in the change in premium, dP, corresponding to a marginal increase in LTCI coverage, dI, that would leave $\mathcal{V}$ unchanged. That is, we focus on a uniform policies. This is consistent with the assumption that k is unobservable to private or public insurers. Analogous to the single-agent case, applying the envelope condition yields

$$\mathcal{WTP} = \frac{\pi_B}{\pi_G} \frac{\int \frac{\partial v_B}{\partial w} dG}{\int \frac{\partial v_G}{\partial w} dG} \quad and \quad \mathcal{MRS} = \frac{\int \frac{\partial v_B}{\partial w} dG}{\int \frac{\partial v_G}{\partial w} dG} \tag{1.3}$$

$\mathcal{WTP}$ and $\mathcal{MRS}$ can be expressed as ratios of the average marginal values of wealth in poor versus good health because the envelope condition implies that (dI, dP) has no extensive margin effect on welfare.[13]

The objective of the rest of our theoretical analysis is to derive an expression for $\mathcal{MRS}$ in terms of responses in c and x to encountering dependency. Our first step is to express $\mathcal{MRS}$ as a function of marginal utilities of food consumption.

[10] This "aggregate" MRS value is the relevant measure for both public and private insurance providers. In our discussion at the end of the paper, we highlight the potential welfare gains from allowing insurance to depend on ex-post consumption adjustments.

[11] If k was observable to the researcher, then the introduction of adjustment costs is trivial since one would simply apply the analyses in the existing literature at each given level of k.

[12] For more general social welfare functions, we require a transformation of G which, in turn, yields different weights for each k. However, it is a priori unclear how normative weights should differ across agents with different adjustment costs.

[13] $\mathcal{MRS}$ coincides with the "average" MRS, $\int MRS(k)dG$, whenever agents in good health do not move. As argued below, this condition is satisfied in most relevant scenarios.

To capture the extensive margin response in housing adjustment, let $\phi_h \equiv \Pr(\{x_h \neq x_0\})$ denote the probability that the agent moves given health, h.[14] To focus on key intuitions, assume that $0 < \phi_B \leq 1$ and $\phi_G = 0$. In Appendix **??**, we show theoretically that this is one of two possible non-trivial cases arising from our model.[15] Given that the vast majority of moving in our data is carried out by households who experience a dependency event, this is also the empirically relevant case.

Lemma 1. *Assume $\phi_G = 0$. Let (c_G, x_G) denote the consumption bundle chosen by agents in good health and (c_B^1, x_B^1) and (c_B^0, x_B^0) denote the consumption bundles chosen by dependent* adjusters *and* non-adjusters *respectively. $\mathcal{MRS}$ can be written as*

$$\mathcal{MRS} = \left[\phi_B \frac{\frac{\partial u_B(c_B^1, x_B^1)}{\partial c}}{\frac{\partial u_G(c_G, x_G)}{\partial c}} + (1-\phi_B)\frac{\frac{\partial u_B(c_B^0, x_B^0)}{\partial c}}{\frac{\partial u_G(c_G, x_G)}{\partial c}}\right] \tag{1.4}$$

Expression 1.4 obtains from (1.3) by using the fact that optimality of food consumption implies that the marginal utility of food consumption in each health-state is equated with the marginal value of wealth in that state. In the standard setting with no adjustment costs, $\phi_B = 1$ and Expression 1.4 reduces to the ratio of marginal utility of food consumption when dependent versus in good health.

However, where there are heterogeneous adjustment costs in housing consumption, Equation 1.4 shows that $\mathcal{MRS}$ is the weighted sum of the respective ratios of marginal utilities of food consumption of *adjusters* and *non-adjusters*. This arises because an agent's decision to move or stay put affects the amount of wealth available for food consumption. This "*liquidity channel*", in turn, implies that the marginal value of insurance differs between agents who move and stay put. As such, since ex-post, the marginal value of insurance differs between adjusters and non-adjusters, a measure of the ex-ante value of insurance would have to weight each case by its probability of occurring.

Equation 1.4 also shows that whenever the marginal utility of food consumption depends on housing consumption, the latter also affects $\mathcal{MRS}$ through a "*complementarity chan-*

[14] As noted earlier, the extensive margin in housing adjustment arises because adjustment costs, $0 < k < \infty$. At the extremes where housing is fully flexible or completely unadjustable ($k = 0$ or $k = \infty$), no such extensive margin exists.

[15] The other non-trivial case has $\phi_B = 0$ and $0 < \phi_G \leq 1$. The trivial case has k sufficiently large that no moving occurs. To be precise, we show that moving under *both* good health and dependency cannot be a solution to the agent's problem whenever (w_1, x_0) is ex-ante optimal. Therefore, assuming that G is such that *some* agents move, we have either $\phi_G = 0, 0 \leq \phi_B < 1$ or $\phi_B = 0, 0 < \phi_G \leq 1$.

nel". Intuitively, when $\frac{\partial^2 u}{\partial c \partial x} > 0$, the same decrease in c has a smaller impact on an agent's welfare if this was paired with a simultaneous decrease in x than if x were unchanged. Since we wish to infer the value of consumption smoothing by using the marginal utility of food consumption, this measure needs to account for the complementarities discussed.

Lastly, Equation 1.4 highlights that if agents consume both flexible and committed consumption goods, then the mapping from the ratio of marginal utilities of consumption to the marginal value of insurance is only valid for *flexible* consumption. Intuitively, because of adjustment costs in housing consumption, changes in the level of housing consumption are determined by both the relative marginal utility of housing consumption across health-states and by the cost of adjustment. Hence, these changes cannot be used to infer the value of smoothing consumption across states. Nevertheless, the *extensive* margin in housing adjustment, ϕ_B, is still relevant as a weight which captures the relative prevalence of each type of housing adjustment (moving versus staying-put).

1.2.3 Consumption-based implementation

The next step in our analysis maps the unobservable marginal utilities of consumption in (1.4) to empirically-observable consumption responses to dependency shocks. This is known in the literature as the "consumption-based approach" (see for example Chetty (2006); Kolsrud et al. (2018)). To facilitate comparison with the canonical consumption-based formula, we assume that preferences over food and consumption are separable, i.e. $\frac{\partial^2 u(c,x)}{\partial x \partial c} = 0$ at each consumption bundle. This abstracts from the "complementarity channel" to focus on the implications of the "liquidity channel". In addition, our baseline result also abstracts from state-dependent preferences by assuming that $u_h(c,x) = u(c,x)$. These restrictions are relaxed successively as extensions below.

Proposition 1. *Assume $\phi_G = 0$ and suppose $\frac{\partial u^2(c,x)}{\partial x \partial c} = 0$ at each (c,x). $\mathcal{MRS}$ can be expressed up to a second-order approximation as*

$$\mathcal{MRS} \approx 1 + \sigma^c \Big[\phi_B \frac{\Delta c^1}{c_G} + (1-\phi_B) \frac{\Delta c^0}{c_G} \Big] + \frac{1}{2} \sigma^c \gamma^c \Big[\phi_B \Big(\frac{\Delta c^1}{c_G} \Big)^2 + (1-\phi_B) \Big(\frac{\Delta c}{c_G} \Big)^2 \Big] \tag{1.5}$$

where $\Delta c^1 \equiv c_G - c_B^1$, $\Delta c^0 \equiv c_G - c_B^0$ are the food consumption responses of adjusters and non-adjusters and ϕ_B is the probability an agent moves when dependent. $\sigma^c \equiv -c_G \frac{\partial^2 u_h(c_G, x_G)}{\partial c^2} \Big/ \frac{\partial u_h(c_G, x_G)}{\partial c}$

and $\gamma^c \equiv -c_G \frac{\partial^3 u_h(c_G,x_G)}{\partial c^3} \Big/ \frac{\partial^2 u_h(c_G,x_G)}{\partial c^2}$ *are the first and second-order measures of the curvature of utility with respect to food consumption, evaluated at* (c_G, x_G)*.*

Proposition 1 expresses $\mathcal{MRS}$ as a function of consumption responses to dependency shocks by applying a set of Taylor approximations. With no adjustment costs (ϕ_B = 1), Equation 1.5 reduces to the canonical consumption-based implementation equation. Broadly, insurance is valuable because agents are averse to variability in their consumption streams. The degree of this aversion is captured mathematically by the curvature in their utility with respect to the focal consumption good. Given this, the consumption-based approach says that at the margin, an agent's marginal valuation of insurance against a shock can be inferred from the magnitude to which the focal consumption good is displaced from its prior optimum due to the shock, scaled by a measure of the curvature in utility with respect to that good.[16]

In the case with adjustment costs in housing, Equation 1.5 highlights two key departures from the standard setting. First, we require separate observations of food and housing consumption responses. Furthermore, when there is no complementarity, we should focus on the *intensive margin* responses in the fully-flexible consumption good, food. Second, we need to distinguish between the consumption responses of adjusters and non-adjusters. These increased informational requirements arise purely from the presence of adjustment costs since we have assumed away state-dependence and complementarity. In fact, state-dependence and complementarity affect the standard implementation formula regardless of the presence of adjustment costs and enter our generalised representation in the same way.

The first implication arises because the consumption-based approach infers the value of consumption-smoothing from local consumption responses to the health shock. Since housing is not adjusted by some agents, only food consumption responses map cleanly, via the scaling term, σ^c to $\mathcal{MRS}$. The second implication arises because adjustment costs induce heterogeneity in the value of insurance between adjusters and non-adjusters via the liquidity channel described above.

Lastly, Expression 1.5 says that we also require an estimate, ϕ, of the extensive margin response in housing to the shock. This is used to weight the respective consumption responses of adjusters and non-adjusters. Note that this extensive margin response is a

[16] The first and second-order measures of curvature in preferences for c are σ_G^c and γ_G^c. In a model where preferences are defined over wealth, σ_G^c and γ_G^c are the coefficients of relative risk aversion and relative prudence.

"sufficient statistic" for capturing the effect of adjustment costs on the marginal value of insurance and obviates the need to impose further assumptions on the distribution of adjustment costs.

Choice of consumption measure and relation to existing studies

We now highlight the implications of violating the two key implications discussed above. This allows us to relate the findings in Proposition 1 to the existing literature using the consumption-based approach.

Neglecting gap between adjusters and non-adjusters

Suppose we observe food and housing consumption separately but do not distinguish between adjusters and non-adjusters. Let this naive measure be denoted $\widehat{MRS}$. Rearranging (1.5) yields

$$MRS \approx 1 + \sigma^c \frac{\Delta \bar{c}}{c_G} + \frac{1}{2}\sigma^c \gamma^c \Big[\alpha\Big(\frac{\Delta c^1}{c_G}\Big)^2 + (1-\alpha)\Big(\frac{\Delta c^0}{c_G}\Big)^2\Big] > 1 + \sigma^c \Delta \frac{\bar{c}}{c_G} + \frac{1}{2}\sigma^c \gamma^c \Big(\frac{\bar{c}}{c_G}\Big)^2 = \widehat{MRS} \tag{1.6}$$

The inequality is a consequence of applying Jensen's inequality to the second-order terms. Thus, neglecting the second implication of committed consumption results in a downward bias of second order magnitude.

Using a measure of total consumption

Let $e_h \equiv c_h + x_h$ denote an agent's total flow consumption, defined as the sum of food and housing consumption. Suppose we only observe total flow consumption, e_h, and not its components. This corresponds to recent applications of the consumption-based implementation which construct broad measures of household consumption.[17] The appropriate measure of MRS is then

$$\begin{aligned} MRS \approx & \phi_B \Big[1 + \sigma^c \Big[\frac{e_G}{c_G}\frac{dc^1}{de}\Big]\frac{\Delta e^1}{e_G} + \frac{1}{2}\sigma^c \gamma^c \Big[\frac{e_G}{c_G}\frac{dc^1}{de}\Big]^2 (\frac{\Delta e^1}{e_G})^2\Big] \\ & + (1-\phi_B)\Big[1 + \sigma^c \Big[\frac{e_G}{c_G}\Big]\frac{\Delta e^0}{e_G} + \frac{1}{2}\sigma^c \gamma^c \Big[\frac{e_G}{c_G}\Big]^2 (\frac{\Delta e^0}{e_G})^2\Big] \end{aligned} \tag{1.7}$$

[17]For example, Kolsrud et al. (2018, 2019); Landais and Spinnewijn (2019) use registry data and Ganong and Noel (2019) use administrative bank account data. A broad measure of household consumption, defined as the sum of food and housing consumption, has also been used in the context of survey data by Meyer and Mok (2019).

Note that we have used $de^0 \equiv dc^0$. Next, since $de^1 \equiv dc^1 + dx^1$, $\frac{dc^1}{de^1} \neq 1$. Thus, the (total) consumption responses of adjusters and non-adjusters require different scaling factors. Specifically, when agents reduce both food and housing due to the shock, the appropriate scaling term for adjusters is $\sigma^c\Big[\frac{e_G}{c_G}\frac{dc^1}{de}\Big] < \sigma^c\frac{e_G}{c_G}$. Similarly, the appropriate second-order scaling term for adjusters is $\gamma^c\Big[\frac{e_G}{c_G}\frac{dc^1}{de}\Big] < \gamma^c\frac{e_G}{c_G}$. Intuitively, since σ^c translates *food* consumption changes to a measure of welfare impact, housing consumption changes in Δe^1 "contaminate" the consumption response estimate and hence the resulting implied MRS.

Expression 1.7 assumes that we have separate consumption response estimates of adjusters and non-adjusters. Now suppose we do not distinguish between adjusters and non-adjusters and the resulting naive measure be denoted by $\widehat{MRS}$.

$$
\begin{aligned}
MRS =& \widehat{MRS} + \phi_B\sigma^c\Big[\frac{e_G}{c_G}\Big]\Big[\frac{dc^1}{de} - 1\Big]\frac{\Delta e^1}{e_G} \\
&+ \frac{1}{2}\sigma^c\gamma^c\Big[\frac{e_G}{c_G}\Big]^2\left(\Big[\frac{dc^1}{de} - 1\Big]\phi_B\Big(\frac{\Delta e^1}{e_G}\Big)^2 + \Big[\phi_B\Big(\frac{\Delta e^1}{e_G}\Big)^2 + (1-\phi_B)\Big(\frac{\Delta e^0}{e_G}\Big)^2\Big] - \Big(\frac{\Delta \bar{e}}{e_G}\Big)^2\right)
\end{aligned}
$$

First ignore the second-order terms. Then to a first-order we have $MRS < \widehat{MRS}$ where the error term is increasing in ϕ_B and decreasing in $\frac{dc^1}{de}$. In line with the preceding discussion, because Δe^1 contains changes in housing consumption, the naive measure, $\widehat{MRS}$ is inflated relative to MRS whenever we use σ^c as the curvature parameter. Turning to the second line, we see two countervailing terms for the second-order error term. The first term corrects for the "contamination" in Δe^1 as described above and is negative since $dc^1 < de^1$. The second term arises from neglecting heterogeneous consumption responses between adjusters and non-adjusters and is positive.

Adjusted curvature coefficient

The preceding discussion showed that different scaling terms are required for adjusters and non-adjusters when only total consumption is observed. Indeed, the fact that consumption commitments induce a difference in curvature coefficients was first noted in Chetty and Szeidl (2007). In the context of the consumption-based approach, this point has typically been used as an argument for adjusting the assumed value of the curvature parameter, σ, whenever consumption commitments are present. However, such adjustments have, thus far, been ad-hoc and not typically guided by theory. We now propose a theoretically-grounded procedure for this adjustment based on our model.

Suppose we have an estimate of the *coefficient of relative risk aversion*, σ^w, which assumes no adjustment costs when they are in fact relevant.[18]

$$\sigma^w \equiv -w \frac{\frac{\partial v^2(w)}{\partial (w)^2}}{\frac{\partial v(w)}{\partial w}} = \sigma^c \frac{w}{c} \frac{d\bar{c}}{dw}$$

Note that we have maintained the assumption of separable preferences. Here, $\frac{d\bar{c}}{dw}$ is the average (over adjusters and non-adjusters) marginal propensity to consume food.

Lemma 2. *Assume $\phi_G = 0$, $\frac{\partial^2 u(c,x)}{\partial x \partial c} = 0$. Furthermore, assume that $\frac{d\bar{c}}{dw} = \phi_B \frac{dc^1}{dw} + (1-\phi_B)\frac{dc^1}{dw}$. Then, $\mathcal{MRS}$ can be expressed to a first-order approximation as*

$$\mathcal{MRS} \approx \sigma^w \Big[\frac{w_G}{e_G} \frac{d\bar{e}}{dw}\Big]^{-1} \frac{\Delta \bar{e}}{e_G} \tag{1.8}$$

The assumption that $\frac{d\bar{c}}{dw} = \phi_B \frac{dc^1}{dw} + (1-\phi_H)\frac{dc^1}{dw}$ says that on average, agents' food and housing consumption responses to a dependency shock match their respective responses to the marginal change in wealth. Recall that any first-order error from using total consumption arises from the "contaminating" effect of including the housing response of adjusters. Since Lemma 2 attempts to map from average (total) consumption changes to an implicit change in wealth and then to welfare measure via σ^w, the assumption ensures that the proportion of adjusters and non-adjusters is unchanged and hence the mapping is valid. This assumption is substantive and may be violated for two reasons. First, it is immediately violated with state-dependent preferences since in that case, changes in health with zero monetary change would affect consumption and welfare and this is not accounted for in σ^w. Second, because housing adjustment choice is endogenous to the size of negative income shock, the pattern of consumption responses to health shock with non-infinitesimal income loss need not be identical to the small wealth perturbations implicit in the definition of σ^w.

Allowing for complementarity in preferences over food and housing

Proposition 1 highlights the implications of the liquidity channel effect of consumption commitments on the consumption-based approach formula. However, with complementarity in preferences over food and housing, the marginal utility of food consumption and its higher derivatives can depend on the level of housing in general. Corollary 1 shows we

[18] For example, this could be taken from studies which estimate σ^w based on broad consumption measures but which do not explicitly account for committed consumption.

can account for this by observing the *intensive margin* response in housing consumption and parameterising the complementarity between the two consumption goods.

Corollary 1. *Assume $\phi_G = 0$. When $\frac{\partial^2 u(c,x)}{\partial c \partial x} \geq 0$, $\mathcal{MRS}$ can be expressed up to a second-order approximation as*

$$\mathcal{MRS} \approx \left\{1 + \sigma^c \Big[\phi_B \frac{\Delta c^1}{c_G} + (1-\phi_B)\frac{\Delta c^0}{c_G}\Big] + \frac{1}{2}\sigma^c \gamma^c \Big[\phi_B \Big(\frac{\Delta c^1}{c_G}\Big)^2 + (1-\phi_B)\Big(\frac{\Delta c^0}{c_G}\Big)^2\Big]\right\}$$
$$- \varepsilon_x^{u_c} \phi_B \frac{\Delta x^1}{x_G} - \frac{1}{2}\varepsilon_c^{u_{xx}} \sigma^x \frac{s_G}{1-s_G}\phi_B \Big(\frac{\Delta x^1}{x_G}\Big)^2 - \varepsilon_x^{u_{cc}} \sigma^c \phi_B \Big(\frac{\Delta c^1}{c_G}\Big)\Big(\frac{\Delta x^1}{x_G}\Big) \tag{1.9}$$

where $\Delta x^1 \equiv x_G - x_B^1$ is the intensive margin *response in housing consumption, $\Delta c^1, \Delta c^0$ are the food consumption responses of adjusters and non-adjusters and $s_G = \frac{q x_G}{c_G + q x_G}$ is housing's share of total consumption expenditure. σ^c and γ^c are the first and second-order measures of the curvature of utility with respect to food consumption and $\sigma^x \equiv -x_G \frac{\partial^2 u_h(c_G, x_G)}{\partial x^2} \Big/ \frac{\partial u_h(c_G,x_G)}{\partial x}$ measure the curvature of utility with respect to housing consumption. In addition, $\varepsilon_x^{u_c} \equiv x_G \frac{\partial^2 u_h(c_G,x_G)}{\partial x \partial c} \Big/ \frac{\partial u_h(c_G,x_G)}{\partial c}$ $\varepsilon_x^{u_{cc}} \equiv x_G \frac{\partial^3 u_h(c_G,x_G)}{\partial x \partial c^2} \Big/ \frac{\partial u_h(c_G,x_G)}{\partial c^2}$, $\varepsilon_c^{u_{xx}} \equiv c_G \frac{\partial^3 u_h(c_G,x_G)}{\partial c \partial x^2} \Big/ \frac{\partial u_h(c_G,x_G)}{\partial x^2}$ are measures of complementarity in preferences over food and housing consumption.*

Furthermore, the complementarity channel has a negative effect on $\mathcal{MRS}$ if

$$\varepsilon_x^{u_c} + \frac{1}{2}\sigma^x \frac{s_G}{1-s_G} \varepsilon_c^{u_{xx}} \frac{\Delta x^1}{x_G} + \sigma^c \varepsilon_x^{u_{cc}} \frac{\Delta c^1}{c_G} \geq 0 \tag{1.10}$$

Since the terms in curly braces in the first line of Equation 1.9 equal Equation 1.5, allowing for complementarity in preferences over food and housing changes $\mathcal{MRS}$ through the terms in the second line. The first term captures a first-order complementarity effect and says that when the marginal utility of food consumption is increasing in housing consumption, the marginal value of insurance would be lower for agents who respond to a dependency shock by decreasing both food and housing. The second-order terms in the second line capture the effect of c on the curvature of preferences with respect to x and the effect of x on the curvature with respect to food. Consolidating these effects, Condition 1.10 says that when increases in c (resp. x) do not reduce the curvature of utility with respect to x (resp. c) by too much (i.e. when second-order complementarity effects are positive or negative but small) the complementarity channel lowers the marginal value of insurance vis-a-vis the case with separable preferences. This would hold, for example, in the case where preferences over c and x satisfy constant elasticity of substitution (CES) with substitution parameter ≥ 1.[19]

Allowing for state-dependent consumption preferences

As noted in Finkelstein et al. (2013), agents' marginal utility consumption arguably differs across health-states. In the context of dependency risk, Ameriks et al. (2016) show that state-dependent preferences over consumption is an important factor for explaining empirically observed LTCI take-up. Given this, to tractably introduce state-dependent preferences into the present framework, consider the simple case where the instantaneous utility function satisfies

$$\frac{\frac{\partial u_B(c,x)}{\partial c}}{\frac{\partial u_G(c,x)}{\partial c}} = \frac{\frac{\partial u_B(c,x)}{\partial x}}{\frac{\partial u_G(c,x)}{\partial x}} = \theta, \qquad \text{for each } (c,x) \tag{1.11}$$

The assumption implies two things: First, dependency affects the agent's marginal utility from consumption of food and housing. Second, it restricts the degree of state-dependence, θ, to be the same for food and housing consumption. The first point helps introduce state-dependence into our setting consistent with past literature. The second point is a necessary implication of the case with non-separable preferences.

Corollary 2. *Assume preferences satisfy Condition 1.11. Assume $\phi_G = 0$. When $\frac{\partial^2 u(c,x)}{\partial c \partial x} \geq 0$, $\mathcal{MRS}$ can be expressed up to a second-order approximation as*

$$\begin{aligned}\mathcal{MRS} \approx \theta \Bigg\{ & 1 + \sigma^c \Big[\phi_B \frac{\Delta c^1}{c_G} + (1-\phi_B)\frac{\Delta c^0}{c_G}\Big] + \frac{1}{2}\sigma^c \gamma^c \Big[\phi_B \Big(\frac{\Delta c^1}{c_G}\Big)^2 + (1-\phi_B)\Big(\frac{\Delta c^0}{c_G}\Big)^2\Big] \\ & - \varepsilon_x^{u_c} \phi_B \frac{\Delta x^1}{x_G} - \frac{1}{2}\varepsilon_c^{u_{xx}} \sigma^x \frac{s_G}{1-s_G} \phi_B \Big(\frac{\Delta x^1}{x_G}\Big)^2 - \varepsilon_x^{u_{cc}} \sigma^c \phi_B \Big(\frac{\Delta c^1}{c_G}\Big)\Big(\frac{\Delta x^1}{x_G}\Big) \Bigg\}\end{aligned} \tag{1.12}$$

Equation 1.12 extends the logic from the previous two results by recognising that with state-dependent preferences, starting from the same consumption bundle, a marginal change in consumption in good health may be valued differently from the same marginal change when in poor health.[20] As such, to mapping from consumption responses to the marginal value of wealth and hence to the marginal value of insurance, requires appropriately weighing the relative importance of (marginal) consumption changes between health states.

[19] See Supplementary Appendix **??**.

[20] As a concrete example, an agent who suffers from long-term care dependency might derive less pleasure from expensive meals or a second-storey extension to their house.

1.3 Empirical application

Do households value the consumption-smoothing ability afforded by long-term care insurance (LTCI)? Our empirical application implements the consumption-based sufficient statistics framework discussed above to address this question in the context of elderly households in the U.S.[21] The goal of our empirical analysis is two-fold: to identify the causal effect of dependency on household flexible and housing consumption and to study heterogeneity in the former across households which do and do not adjust their housing. The first objective, which seeks to estimate the empirical counterpart of $\Delta c/c$ from our theoretical model, provides a baseline answer to our question. In turn, studying the interaction between responses in flexible and committed consumption qualifies this baseline answer and highlights the implications consumption commitments on the implied value of LTCI. In this section, we briefly outline our data, set out key definitions and describe our strategy for this empirical analysis.

1.3.1 Data and institutional context

Long-term care insurance in the U.S. Long-term care (LTC) expenditures form a significant proportion of medical spending amongst the elderly. This burden increases towards the end of life, with LTC accounting for around 45 percent of total medical spending in the last three years of life (French et al., 2018). Despite this, LTC expenses are not covered universally through Medicare. Furthermore, private LTCI ownership rates are also low, at around 10 percent in 2008, and skewed towards the higher end of the wealth distribution (Lockwood, 2018). On average, a third of LTC spending in the U.S. is out-of-pocket, with about 60 percent paid through means-tested public insurance via Medicaid and the remaining (about 4 percent) covered by private insurance (Brown and Finkelstein, 2011). Within this institutional context, our application focusses on nursing home episodes as our focal long-term care dependency event and seeks to infer the marginal value of insurance against these events.

The Health and Retirement Study Our analysis uses data from the Health and Retirement Study (HRS).[22] The HRS is a biennial longitudinal survey of a representative sample of the U.S. population over 50 years old. Starting from Wave 1 in 1992, sampled individ-

[21] That is, we implement Equation 1.5 by estimating empirical counterparts of $\Delta c/c$'s and ϕ.

[22] The HRS (Health and Retirement Study) is sponsored by the National Institute on Aging (grant number NIA U01AG009740) and is conducted by the University of Michigan.

uals and their spouses are interviewed at approximately two-year intervals. Each interview collects information on income, wealth and food expenditure at the household level along with individual-level information on demographics and health care utilisation and expenditure for the main sampled individual and their spouse. We use all available data from Waves 1 to 13 (1992-2016) of the HRS and we index calendar time with these survey wave numbers, i.e. $t = 1, 2, \ldots, 13$.

Our unit of analysis is a household. Households in the HRS can contain one or more residents for whom we have individual-level information on up to two respondents (main sampled individual and their spouse, if any). On average, households in the HRS contain 2.2 residents with around 61.5 percent of households composed of married/partnered couples. For concreteness, our discussion often refers to the running example of a household consisting of two spouses. With respect to this example, the aim of our empirical work is to study how household consumption responds to one of the spouses becoming dependent.

Definition of dependency We operationalise "dependency" by focussing on nursing home stays. At each interview, the HRS collects information on the number of times (if any) a respondent or their spouse has stayed at a nursing home, the total number of nights spent in care and whether or not that individual is residing in the nursing home at the time of interview. Out-of-pocket expenditures associated with nursing home stays are also reported.

We define a "nursing home event" (NH-event) as a household's first observed nursing home stay lasting at least 90 nights. For cases where different spouses from the same household enter a nursing home at different times, we consider only the first of those qualifying stays. This threshold reflects the average elimination period of private LTCI contracts and also addresses the fact that Medicare covers the cost of skilled rehabilitative care up to 100 days.[23] To distinguish between individuals, we shall hereafter refer to the spouse who enters the nursing home as the "dependent spouse". With this definition, the empirical results in subsequent sections can be interpreted as pertaining to the implied value of an insurance contract which covers the expenses associated with nursing home use and which has a 90-day elimination period.

As noted by Hurd et al. (2014), two important components of nursing home use is the (lifetime) risk of having a nursing home stay and the duration of these stays. To capture

[23] Since rehabilitative care cases form a (small) subset of nursing home cases, this cut-off is more stringent than implied by the actual institutional context.

the latter, we construct variables on nursing home *spells* and the duration of these spells by using the panel structure of the HRS and using the diverse information on nursing home stays collected during interviews.[24] Constructing these variables allows us to not only identify instances of nursing home *use* but also know if these correspond to spells which extend across multiple survey waves.[25]

Definition of flexible consumption We take household food consumption as our main measure of flexible consumption. Household expenditure on food is defined as the sum of reported expenditures on food at home, food outside of the home and the value of food stamps received.[26] For our analysis, we assume that such expenditures provide a valid measure of actual food consumption and, with some abuse of terminology, use "food consumption" to refer to expenditure on food.[27]

Definition of committed consumption We focus on housing as our committed consumption good. For our baseline analysis, we assume that preferences over food and housing consumption are separable. From Equation (1.5), this implies that we need only focus on extensive margin responses in housing consumption.

We take reports of home sales as our primary measure of extensive margin housing adjustment. At each wave, the households report if they have made any sale or purchase of a primary residence or second home since the last interview. Of the reported transactions, 77.5 percent involve households selling only or selling and buying their home. Since we restrict our analysis sample to households who are homeowners prior to encountering a nursing home event (details in the next subsection) and since the large majority of all households in the HRS own only one home, this measure matches its theoretical counterpart of "moving" quite closely.

As a robustness check, we also repeat our analysis with reported "home-ownership status" changes as an alternative definition of an extensive margin housing adjustment. This alternative measure is constructed by comparing a household's home-ownership status at each wave with their status in the immediately previous wave. Given our analysis sample, a change in home-ownership status after onset of the nursing home event implies that the household has transitioned from home-ownership to renting.[28] This is a significantly

[24] See Data Appendix for details

[25] Conversely, it also allows us to separate nursing home spells corresponding to the focal "nursing home event" from subsequent re-entries into a nursing home.

[26] See data appendix for details on construction of this variable from raw HRS data.

[27] This implicitly assumes that all expenditures are consumed and that home-production of food can be neglected.

[28] Where "renting" is taken to include living in a rented residence and living with relatives/friends/etc.

narrower definition as it does not account for "downgrading", whereby households sell off a previously owned home and then purchase a new residence.

1.3.2 Empirical strategy

To study the effect of a nursing home stay on household outcomes, we adopt the following model of dynamic treatment effects. Let i index households, t index calendar time and suppose household-i encounters a nursing home event at time $t = \tau$. Let $j \equiv t - \tau$ denote event time, i.e. the time between survey wave, t, and the wave in which the NH-event is encountered ("event-time 0"). We assume that an outcome of interest, Y_{it}, can be described by:

$$Y_{it} = \alpha_i + \gamma_t + \sum_{j=\underline{J}}^{\overline{J}} \beta_j \times \mathrm{I}_j + \zeta X_{it} + \varepsilon_{it} \tag{1.13}$$

Here, α_i captures time-invariant household characteristics and γ_t is the effect of calendar time. I_j's are indicators for event time-j and $[\underline{J}, \overline{J}]$ is a window around nursing home entry wave. Our objective is to identify the set of dynamic effects $\{\beta_j\}_{j=\underline{J}}^{\overline{J}}$ of the NH-event on Y_{it}.

In line with the recent literature, we estimate Specification 1.13 by fixed-effects regression on a sample of households with observed NH-events (treatment group) and a matched control group composed of households which never report any nursing home use in all waves they are observed. The inclusion of "never treated" households allows for separate identification of time-varying effects which might otherwise be confounded with the dynamic treatment effects of interest.[29]

The identifying assumption is that absent realisation of the nursing home event, consumption of treated and control households follow the same trajectory. The credibility of this parallel trends assumption relies on the comparability of control and treated households. To this end, we use nearest-neighbour matching to select our control households.

Specifically, at each survey wave, t, we take treated households who first encounter their nursing home event at t and match them to households in the never-treated pool us-

[29]Figure 1.15 plots annual food expenditure against age and shows that a declining consumption-age profile is one of the possible time-varying confounders.

ing characteristics at pre-event waves.[30] We match exactly on the dependent spouse's age and by propensity score on gender, education level, race, partnered status, census division location, number of household residents, number of living children, total non-housing wealth, value of primary residence, three lags of total household income and total out-of-pocket medical spending and the dependent spouse's actuarial long-term care risk score.[31] Controlling for a pre-event measure of risk is a succinct way of accounting for the effect of underlying NH risk factors which may affect both household consumption and nursing home utilisation.

Accounting for the mechanical change in household size

Changes in household size are particularly relevant confounders as, by definition, an individual who stays in a nursing home does not consume within their own home. Given this, consumption within the household is expected to decrease mechanically during to a nursing home stay. However, this effect does not correspond to the behavioural response that we seek. As such, ignoring the mechanical effect due to a change in household size associated with a NH-event would overstate the magnitude of a consumption drop in response to a NH-event and, subsequently, bias upwards the implied marginal value of insurance.

To address this one could adjust for the number of residents in the household at each HRS interview. However, because the number of residents in a household is elicited at the time of an interview, the this approach understates the change in the number of residents over the period between two waves whenever entry into and exit from a nursing home occurs entirely between interviews. In contrast, this approach overstates the change in the number of residents if the spouse residing in the nursing home at the time of interview has not spent the entire duration between survey waves in the nursing home.[32]

To address this issue, we first construct a measure of "effective household-size" using information on the duration of nursing home spells. Specifically, define η_{it} to be the proportion of time a dependent spouse of household-i spends in a nursing home, as reported at interview-t.[33] The effective household size accounts for three different cases: first, for households with no nursing home entry, the effective household size is equal to the re-

[30] See Table 1.6 for details.

[31] This objective measure of dependency risk is computed for each individual using the actuarial model of **?**.

[32] This is case for *all* households whose dependent spouse is residing in the nursing home at the $j = 0$ interview because, by construction, *no* respondents are in a nursing home at $j = -1$.

[33] In practice, η_{it} is the number of nights spent in a nursing home between t and $t-1$ divided by the number of days between interviews t and $t-1$.

ported household size. Next, for households reporting a nursing home entry but whose dependent spouse is *not* resident in the nursing home at the time of interview, the effective size is defined as the reported household size minus η_{it}. Finally, for households whose dependent spouse *is* resident in the nursing home at the time of interview, we define the effective change as the reported household size plus $1 - \eta_{it}$.

Using this measure of the effective household size, we next define the effective change in household size with respect to its pre-event size as the difference between the effective household size at event-time $j \neq -1$ and the household size at $j = -1$. i.e. $\Delta HHRes_{i,j} \equiv HHRes_{i,j} - HHRes_{i,0}$.[34] Finally, to account for the mechanical effect of nursing home episodes on household consumption via the change in household size, we include as covariates the interactions between effective change in household size and indicators for the number of residents at $j = -1$, i.e. $1\{HHRes_{i,0} = n\} \times \Delta HHRes_{i,j}$ for $n \in \{1, 2, 3, 4$ and above$\}$ and the squared interactions $1\{HHRes_{i,0} = n\} \times (\Delta HHRes_{i,j})^2$. These interactions allow for the inter-temporal effect of a change in household size to differ across households of different initial size.[35]

Estimation sample

Since the marginal value of insurance can be expected to differ depending on one's existing insurance coverage, we restrict our analysis to households who have no pre-existing LTCI coverage.[36] This implies that our analysis can be interpreted as inferring the marginal value of providing insurance to the uninsured.

Next, since we take observations of home sales to be our definition of an extensive margin housing adjustment, we also restrict our analysis sample to households who report being home-owners at pre-event waves. As adjustment costs associated with moving is likely to differ systematically between homeowners and renters, this restriction also helps to account for the confounding effect of home-ownership status on the probability of adjusting

[34]We use the event-time index here to emphasise that the change is defined with respect to the pre-event wave, $j = -1$.

[35]The distinction between cross-sectional and inter-temporal differences in household size is particularly relevant for committed consumption such as housing. As a stylised example, let housing consumption be defined as the number of bedrooms in a house and suppose each resident in a household requires one bedroom. Then the cross-sectional difference in housing consumption between a two-member and three-member household is one. Now, focus on the three-member household (without loss) and consider the departure of one of its members. This registers as a one-unit difference in the number of residents across time, but *need not* equate with a one-unit change in housing consumption if, for example, the previously occupied bedroom is left vacant.

[36]Here, we define LTCI coverage as either having private LTCI or being enrolled in Medicaid. This restriction implies we, restrict analysis to the subset of "treated" households who do not have LTCI at pre-event waves $j < 0$ and match them to a pool of candidate control households who are similarly uninsured.

housing. Unfortunately, this restriction comes at the cost of generalising our results to the wider population.

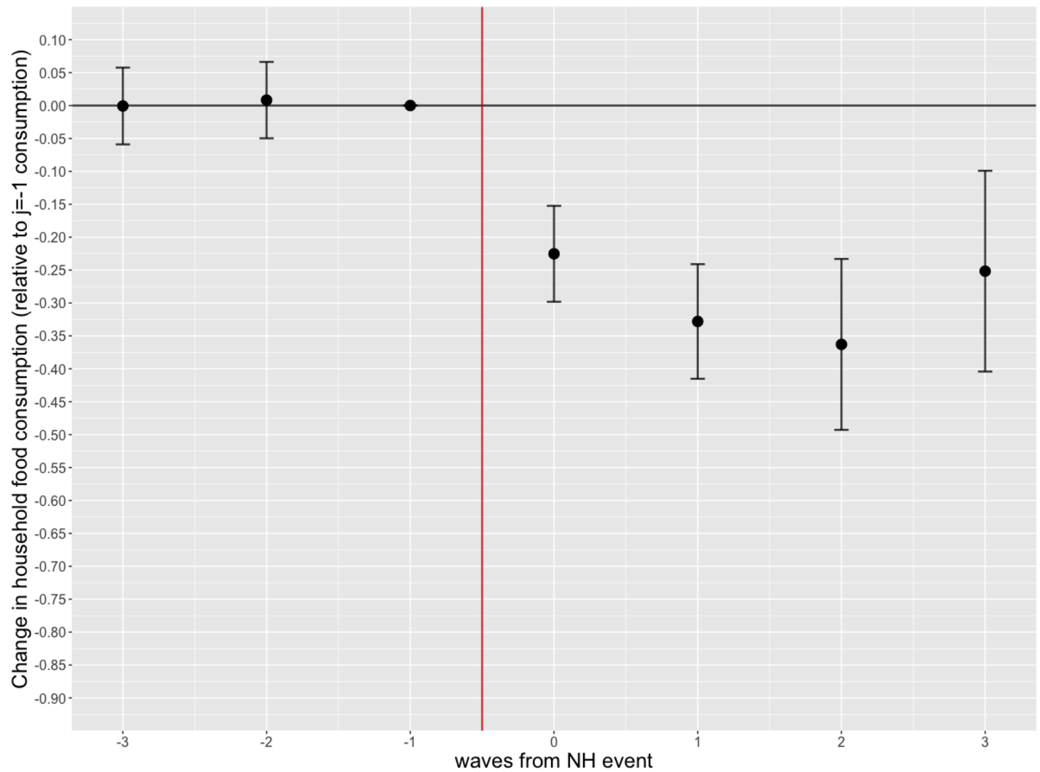
Descriptive statistics on key characteristics of our analysis sample are displayed in Table 1.1 . Overall, pre-exclusion and post-matching samples are largely similar in their demographic characteristics. However, due to exclusion of pre-existing Medicaid enrollees, households post-exclusion have, on average, higher total net assets and total income. The latter group's higher average wealth level derives both from higher levels of total non-housing wealth along with higher-valued primary residences.

1.4 Response in flexible consumption to nursing home stays

In this section, we focus on the first objective of our empirical analysis: to estimate the response in households' flexible consumption to a nursing home stay. As preliminary evidence, Figure 1.7 plots the mean annual food expenditure of households in our treatment (in red) and matched control groups (in blue) for a window spanning three survey waves before and after the initial entry wave. Visually, three relevant features stand out: first, food consumption of both groups at pre-event wave $j = -3, -2, -1$ follow largely identical trends and are indistinguishable in levels as well. Second, average food consumption amongst treated households drops sharply at $j = 0$, the wave in which the nursing home episode commences, while a similar decrease does not occur for control households. Third, consumption continues to decrease from $j = 0$ to $j = 1$ before levelling off and remaining at a persistently low level relative to consumption of control households. The evidence of similar pre-event trajectories provides support towards our assumption of parallel trends between treatment and matched control households while the sharp deviation from this path upon nursing home entry suggests a significant consumption response to this event.

To quantify these effects, we estimate Specification 1.13 using households' reported annual food expenditure as the dependent variable. Taking $j = -1$ as the reference, Figure 1.1 plots the estimates of β_j for $j = -3, \ldots, 3$ normalised by the mean annual consumption of treated households at $j = -1$, i.e. $\beta_j / \hat{C}_{-1}$. These estimates condition on our full set of demographic and financial covariates along with controls for the effective change in household residents, as described above. Formalising our initial observations, the figure shows that households respond to a spouse's nursing home entry by reducing annual food consumption by about 23 percent of pre-event levels in the immediate nursing home entry wave ($j = 0$). Reported annual consumption decreases further at $j = 1, 2$ to around 32 percent below pre-event levels before recovering slightly at $j = 3$.

Figure 1.1: Dynamics of household food consumption around NH entry



Note: The figure plots the estimates from specification 1.13 with household annual food consumption as the dependent variable. Each point represents β_j, for $j = -3, \ldots, 3$ normalised by average consumption of the treatment group at $j = -1$. These estimates are control for the effective change in household size, as described in the main text, and on the full set of demographic and financial covariates given in Table 1.5. 95 % confidence intervals are based on cluster robust standard errors with clustering at the household level.

How are these estimates affected by our control for the mechanical change in household size? Figure 1.8 plots the same normalised treatment effect estimates, under three different covariate specifications. As expected, omitting household size controls (in empty squares), yields consumption response estimates that are significantly inflated at all post-event times. In contrast, using only information on the reported change in number of household residents at each wave (grey squares) tends to overstate the *mechanical reduction* in household size and hence underestimate the consumption *response* to a nursing home episode. This is especially evident at nursing home entry wave since by construction, treated respondents could not have spent the entire duration between waves in the nursing home.

The fact that estimates using the effective change in household size converges towards those using reported changes suggest that some of the persistent consumption decrease at later event-waves may be due to respondents whose spells continue into these later periods. Figure 1.14 explores this fact using two alternative ways of representing the length respondents' nursing home spells. Panel 1 plots the distribution of nights spent in a nursing home during respondents' index nursing home spell and shows that while a significant portion of nursing home spells are shorter than 720 nights (roughly the time between interviews), there is a significant right skew in the distribution of index spell durations. Panel 2 of Figure 1.14 provides a complementary representation in terms of event-time and shows that of more than half of all index nursing home spells continue into event-wave $j = 2$ and more than 40 percent last beyond $j = 3$. Together, these figures reiterate that the use of care *and* the duration of such care use are both important components of long-term care risk.

1.4.1 Dynamics of medical expenditure and income

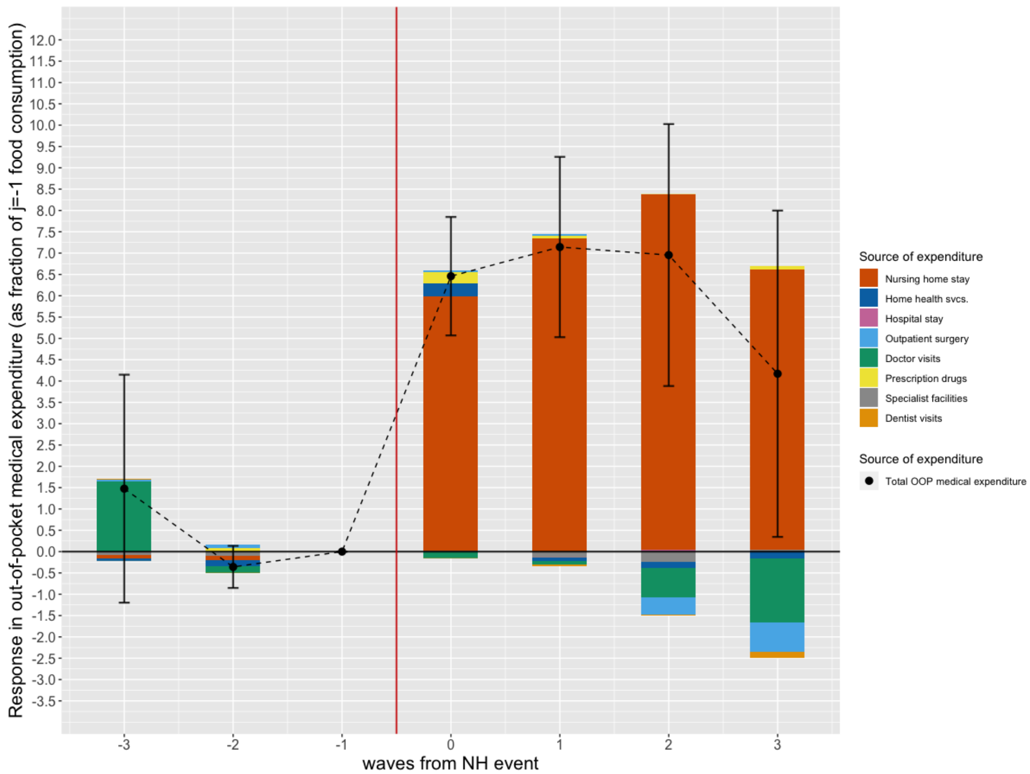
Are the decreases in consumption documented above a *response* in household consumption to the shock associated with a nursing home episode and not vice-versa? To probe into this, we study the dynamics of out-of-pocket (OOP) medical expenditures and household income to the focal nursing home event by repeating our event-study analysis with various dependent variables from these two categories.

Figure 1.2 shows the dynamics in total OOP medical expenditure to the nursing home event and breaks this down into its component sources of expenditure.[37] Each point represents the estimates of β_j for $j = -3, \ldots, 3$ normalised by the mean household food expenditure of treated households at $j = -1$. The figure shows that nursing home entry is associated with sharp increase in OOP medical expenditures of 6.5 times the amount of pre-event annual food spending. This increase is driven predominantly by OOP nursing home expenses and persists into the two subsequent waves before decreasing at $j = 3$.

In stark contrast to the large and statistically significant increase in out-of-pocket medical expenditures, Figure 1.9 shows that the response in household incomes is measured very imprecisely. This is driven in part by the fact that recipient-ship of income from these sources is imbalanced amongst households. Imprecision aside, the estimates also suggest that changes in household income around the nursing home event are also of much

[37]By definition, total OOP medical expenditure is the sum of OOP spending on nursing home stays, hospital stays, outpatient surgery, doctor visits, prescription drugs, home health care services, specialist healthcare facilities and dentist visits. To be precise, we estimate Specification 1.13 for Y_{it} defined as each of the above expenditure categories.

Figure 1.2: Change in out-of-pocket medical expenditures around NH entry



Note: The figure plots the treatment effect estimates from specification 1.13 for total out-of-pocket medical expenditure and its constituent categories for $j = 0, \ldots, 3$. All estimates are expressed as a percentage of average consumption of the treatment group at $j = -1$. 95 % confidence intervals for the change in total OOP expenditures are based on cluster robust standard errors with clustering at the household level.

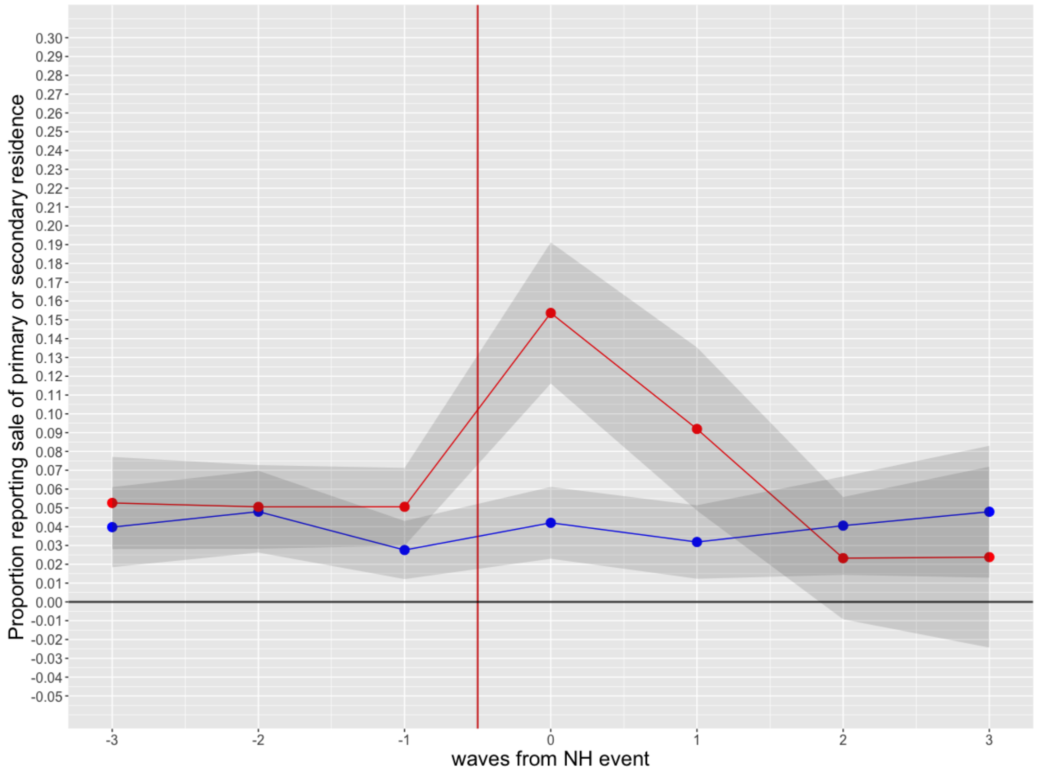
smaller magnitude relative to the corresponding expenditure increases.

Together, the findings from both figures suggest that households encountering a nursing home spell face a net negative income shock. This is consistent with how we modelled dependency theoretically and suggests that part of the decrease in household food consumption is a response to this income shock.

1.5 Extensive margin response in housing consumption

As with food consumption, households could also to respond to a spouse's nursing home encounter by adjusting their housing consumption. However, with adjustment costs, not all households would necessarily make this adjustment. To understand households' extensive margin response in housing consumption, that is, their decision of whether to adjust housing consumption, we now turn to study the dynamics of home sales around a spouse's nursing home entry.

Figure 1.3: Proportion of households reporting home sales in treatment and control groups

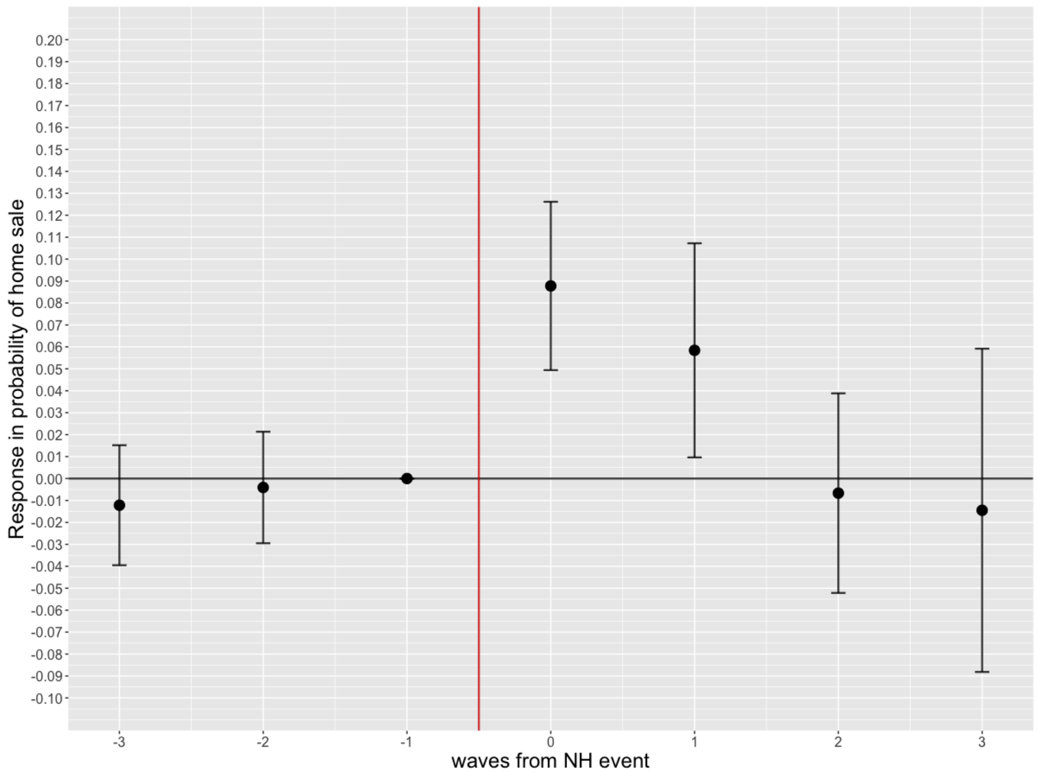


Note: The figure plots the proportion of households reporting a sale of a primary residence or second home at event times $j = -3, \ldots, 3$ for treated (red) and control groups (blue). The grey bands depict 95% confidence intervals for each mean.

Figure 1.11 provides preliminary evidence by plotting the proportion of treatment (red) and control (blue) households which report selling their primary residence or second home for the window spanning three waves before and after the nursing home entry wave. Overall, households in both groups make such extensive margin housing adjustments very infrequently. Specifically, only about five percent of households reporting a home sale in any pre-event wave. As with food consumption, we also see that the probability of home sales amongst treated and control households follow largely similar trends at $j = -3, -2, -1$. This pattern is broken at the nursing home entry wave, where the probability of a home sale increases sharply amongst treated households.

To study this response more formally, Figure 1.16 plots the estimates of β_j from Specification 1.13 where the dependent variable is an indicator for a reported sale of a household's primary or secondary residence. It shows the probability of making this form of housing adjustment increases by about 9 percentage points at the immediate onset of a nursing home episode. The effect of the nursing home event remains positive at $j = 1$ before

Figure 1.4: Dynamics of probability of home-sales around NH entry



Note: The figure plots the estimates from specification 1.13 an indicator for a reported sale of a primary residence or second home as the dependent variable. These estimates are control for the effective change in household size, as described in the main text, and on the full set of demographic and financial covariates given in Table 1.5. 95 % confidence intervals are based on cluster robust standard errors with clustering at the household level.

returning to zero at later times.

1.5.1 Dynamics of implied housing consumption

Economic intuition suggests that households who do make costly housing adjustments during a nursing home episode do so by reducing housing consumption. To explore if this intuition is empirically verified, we study two further measures of housing consumption.

First, Figure 1.16 plots the estimates of β_j from Specification 1.13 where the dependent variable is an indicator for a change in home-ownership status. Overall, the dynamics of the probability of home-ownership status changes mirrors that of the probability of home sales. However, the estimated effect sizes are almost twice the size in the case of home-ownership status changes. This supplemental analysis suggests two points: first that households do indeed respond to the shock associated with a nursing home episode by reducing housing consumption. Second, while transitions out of home-ownership are, on average, very rare, nursing home events can be sufficiently severe for some households to make such a large adjustment.

Next, given that some households respond to a spouse's nursing home episode by moving out of their previous homes, to what degree do they adjust their housing consumption? To provide some indicative evidence regarding the intensive margin response in housing consumption, we take 8 percent of the reported value of one's primary residence as a proxy measure of a household's annual housing consumption (this follows Meyer and Mok (2019)). Using this measure as the dependent variable in our event-study setup yields the estimates of the intensive margin housing response displayed in Figure 1.17. These estimates imply that households which do make adjustments to housing consumption do so by "downgrading" to houses of lesser value. Despite this suggestive evidence, caution needs to be taken in relating these estimates to earlier findings regarding extensive margin responses since the present intensive margin responses are derived from households who have not transitioned from home-ownership to renting.

1.6 Implications for the consumption smoothing value of LTCI

Estimating flow consumption responses

The baseline specification yields estimates of the effect of an NH-event on consumption as reported at each wave, t. However, implementation of the consumption-based measures of the marginal value of LTCI requires estimates of flow consumption changes in response to the NH-event. To obtain the latter, we consider the following specification

$$Y_{it} = \alpha_i + \gamma_t + \sum_{j=-3;j\neq 0}^{3} \beta_j \times \mathrm{I}_j + \beta_0 \times \eta_0 \times \mathrm{I}_0 + X_{it}\zeta + \varepsilon_{it} \tag{1.14}$$

Equation 1.14 differs from Specification 1.13 in that it multiplies I_0, the indicator for the initial NH entry wave, with η_0, the proportion of time in NH between $j = 0$ and $j = -1$. This adapts the simple parametric approach in **?** for identifying flow consumption changes. Intuitively, if the drop in households' food consumption over a reference period is linear in the duration a respondent is resident in the nursing home during that reference period, then the flow consumption decrease is given by the product $\eta_0 \times \mathrm{I}_0$. Here, identification relies on two assumptions: first consumption responses are zero prior to $j = 0$ and second, the decrease in reported annual food consumption scales linearly in the amount of time spent in the nursing home over a reference period. To provide visual support for the second assumption, Figure 1.12 plots the estimated decrease in annual food consumption against the elapsed time a respondent has spent in a nursing home at each interview.[38] The figure shows that the decrease in reported annual food consumption is approximately linear in the length of time respondents spend in nursing home. Given that nursing home expenses typically involve constant residency fees, this finding is consistent with our intuition regarding the structure of nursing home expense risk.

Given the preliminary evidence, the estimates of β_0 from running Specification 1.14 using our four different sets of covariates are displayed in Columns 1 to 4 of Panel 1 of Table 1.4. Dividing these estimates by the mean food consumption of treated households at $j = -1$ yields the corresponding flow consumption responses in Columns 1 to 4 of Panel 2. Focussing on Column 4, which includes our preferred set of control covariates, shows that on average, households experience a 26.6 percent flow consumption decrease upon encountering a nursing home spell. To illustrate how these flow consumption response estimates map to the consumption-smoothing value of LTCI, Columns 1 to 4 in Panel 3 display the implied MRS under the assumption of $\sigma^c = 1$. These estimates are obtained by substituting the flow consumption response estimates and $\sigma^c = 1$ into the naive implementation formula **??** because this baseline specification does not account for different consumption responses across the extensive margin for housing adjustment. Here, focussing on Column 4 yields an implied "naive" MRS estimate of 1.34 which says that uninsured households in our sample are willing to pay around a 34 percent markup over the actuarially-fair rate for insurance.

Allowing for heterogeneous consumption responses

[38] To obtain these estimates, we first restrict the sample to the subset of treated households whose respondent is resident in a nursing home at the time of each interview. At each event-wave, j, we then define an indicator, $D_{i,j}$, for whether the respondent has spent ≤ 365 nights in NH between $j-1$ and j. This allows us to define, at each event-wave, respondents who have spent up to one year in NH and those who have spent (approximately) up to two years in NH. We then estimate the augmented specification

$$Y_{it} = \alpha_i + \gamma_t + \sum_{j=-3}^{3} \beta_j \times \mathrm{I}_j + \sum_{j=-3}^{3} \pi_j \times D_{i,j} \times \mathrm{I}_j + X_{it}\zeta + \varepsilon_{it} \tag{1.15}$$

Given this, π_0 is the decrease in annual food consumption of respondents spending up to one year in NH, β_0 the decrease in annual food consumption for those spending up to two years in NH, π_1 the corresponding decrease for those spending up to three years in NH, β_1 the response for those spending up to four years in NH and similarly for the remainder of the estimates.

Given the above "naive" baseline estimates, we next turn to study heterogeneity in consumption responses between households which adjust their housing in response to a nursing home episode versus those who do not. To do so, we define "Movers" to be treated households which are observed to have sold their homes at $j = 0$ and "Non-movers" as the complementary group. i.e. Movers are households which make the extensive margin housing adjustment, as defined and studied in the Section 1.5. Our focus on adjustment activity at $j = 0$ is driven by our observation that most of the extensive margin housing response takes place at $j = 0$ and the fact that a significant proportion of dependent respondents are still resident in the nursing home at the $j = 0$ interview but this proportion decreases in subsequent periods.[39] With this definition, we consider the following augmented specification

$$\begin{aligned} Y_{it} =& \alpha_i + \gamma_t + \sum_{j=-3;j\neq 0}^{3} \beta_j \times \mathrm{I}_j + \sum_{j=-3;;j\neq 0}^{3} \delta_j \times \mathrm{Move}_{i,j} \times \mathrm{I}_j \\ &+ \beta_0 \times \eta_0 \times \mathrm{I}_0 + \delta_0 \times \mathrm{Move}_{i,j} \times \eta_0 \times \mathrm{I}_0 + X_{it}\zeta + \varepsilon_{it} \end{aligned} \tag{1.16}$$

Specification 1.16 differs from 1.14 by including the interaction terms $\mathrm{Move}_{i,j} \times \mathrm{I}_j$, where $\mathrm{Move}_{i,0}$ is equal to one whenever household i reports a reported home sale at $j = 0$. This allows for differential consumption responses at all j between Movers and Non-movers and allows us to pin down the difference in flow consumption responses by examining β_0 and δ_0. In particular, the flow consumption responses of Non-movers is given by β_0/c_{-1} while the flow consumption responses of Movers is given by $\beta_0/c_{-1}+\delta_0/c_{-1}$. Here, a formal test of heterogeneity in consumption responses between Movers and Non-movers is a test of the null hypothesis $H_0 : \delta_0 = 0$. In addition, an implication of the liquidity channel, as expressed in Lemma **??**, is that δ_0 is strictly positive. Formally, this can be tested with the null $H_0 : \delta_0 \leq 0$.

To provide exploratory graphical evidence of how consumption responses might differ between our two groups of interest, Figure 1.13 repeats the plot from Figure 1.12 separately for Movers (blue) and Non-movers (red). Recall that the calendar time between survey waves is approximately two years. Thus, Movers are households which report selling their homes at some point within the first two years of their nursing home spell. Given this, Figure 1.13 is supports the intuition of the liquidity channel by showing that the consumption drop during a nursing home spell is substantially attenuated when af-

[39] Recall from our theoretical analysis that we are interested in the difference in flow consumption responses *within* their dependency spell.

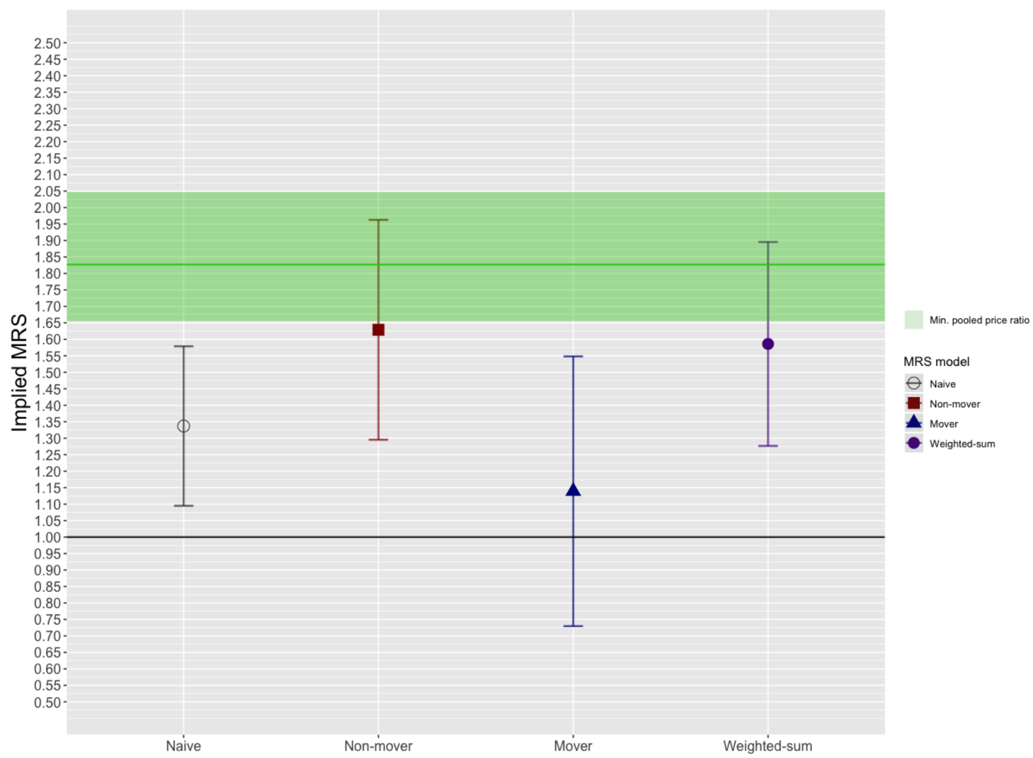
fected households adjust their housing. This difference is most evident at Year-2, which corresponds to the immediate two-year window after a household sells their home but returns to zero as respondents spend more time in the nursing home.

Building on this indicative evidence, Columns 5 to 8 of Panel 1 of Table 1.4 give the estimates of β_0 and δ_0 from Specification 1.16. Focussing on Column 8, which accounts for our preferred set of control covariates, the estimates in Panel 1 show that δ_0 is positive but only statistically-significant at the 10% level. Nevertheless, working with these estimates and following the same steps as in the previous section yields implied flow consumption drops of 43.8 percent for Non-movers and a corresponding drop of 12.4 percent for Movers. This reiterates the graphical finding from Figure 1.13 that households who adjust housing in response to a nursing home episode tend to reduce food consumption to a lesser degree compared to households who do not adjust their housing consumption. As shown in Panel 3, this difference in flow consumption response translates to a gap in the marginal value of insurance between the two groups. In particular, with $\sigma^c = 1$, the resulting implied MRS's suggest that while Non-movers are willing to pay around a 63 percent markup over the actuarially-fair rate for insurance, Movers are only willing to pay a 14 percent markup.

Implied marginal value of insurance
The implied MRS obtained above are conditional on households' ex-post response in housing consumption. However, to the extent that insurance cannot be conditioned on ex-post consumption responses of enrollees, a relevant measure of the value of insurance would still need "average over" the corresponding values of these two groups. This can be done by implementing our weighted sum implementation formula, 1.5. To compare the result of this approach to our earlier "naive" implementations, Figure 1.5 plots the implied MRS from our weighted sum measure under $\sigma^c = 1$ along with the MRS estimates obtained from the analysis in the previous two sub-sections. A number of points are worth remarking here: first, note that the estimates from both the naive and weighted sum implementations are bounded from above by the MRS of Non-movers and from below by the MRS of Movers. This is intuitive since both measures are effectively inferring an average marginal value for insurance between these two groups. Second, comparing the naive against the weighted sum measure shows that the latter yields a slightly higher estimate. This is consistent with the result in Corollary **??** which says that using Equation **??** to infer the MRS for insurance in a setting where consumption commitments are relevant results in an underestimate which is second-order in magnitude.

To contextualise our findings, Figure 1.5 plots the minimum pooled price ratio for rejectees in the market for LTCI (Hendren, 2013) as the green line (and confidence interval as the shaded green band). In our context, this estimate can be interpreted as the minimum markup above the actuarially-fair rate that the uninsured in our sample must be willing to pay in order for trade to occur in a private LTCI market. Broadly speaking, this estimate can be seen as the asymmetric information cost of providing insurance in our present setting. For a given level of risk aversion, comparing this cost against the marginal utility benefit as implied by our MRS measures thus provides an indication of whether or not we would expect trade in a private market for LTCI. In this respect, Figure 1.5 suggests that under the assumption of $\sigma^c = 1$, no trade would be expected.

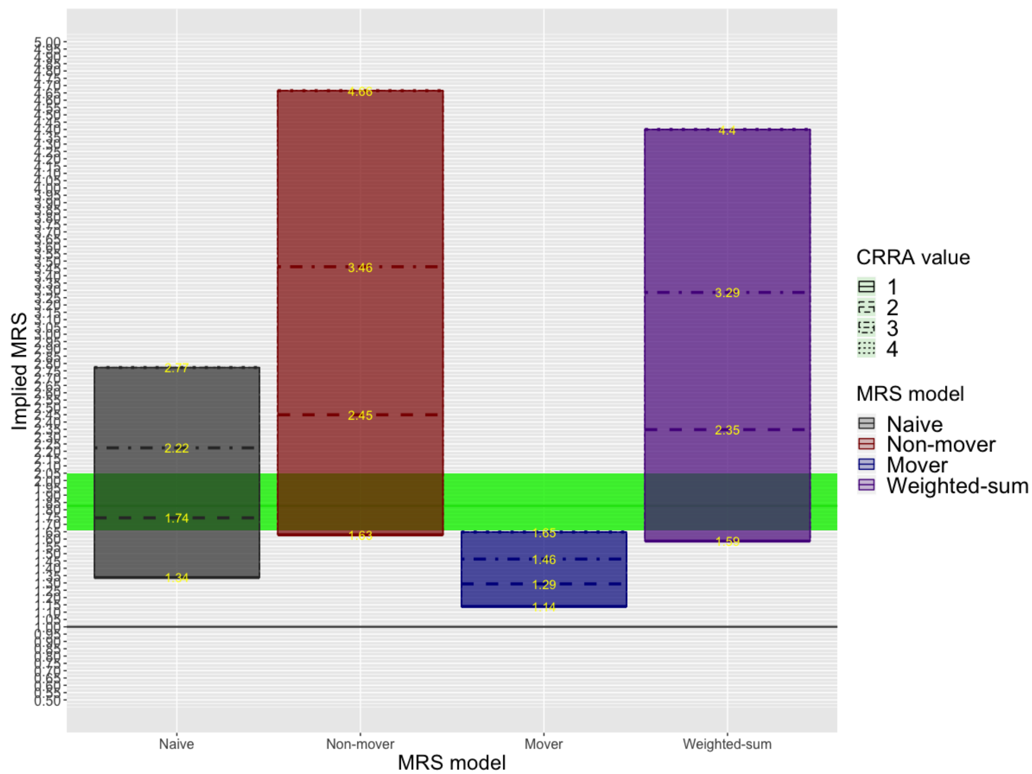
Figure 1.5: Implied MRS estimates from naive and weighted-sum implementations



Note: The figure plots the implied MRS under the assumption of $\sigma^c = 1$ under various implementation approaches. The empty circle corresponds to the naive implied MRS from substituting the average flow consumption response from Column 4 of Table 1.4 into Equation **??**. The red square and blue triangle plot the group-specific implied MRS' obtained from substituting the respective flow consumption responses of Non-movers and Movers (Column 8 of Table 1.4) into Equation **??**. The purple circle represents the implied MRS from using substituting the respective flow consumption responses of Non-movers and Movers together with the probability of home sale at $j = 0$ into the weighted-sum implementation formula 1.5. 95 % confidence intervals are computed by the Delta method and based on cluster robust standard errors with clustering at the household level. The green line and corresponding confidence bands depict the minimum pooled price ratio for rejected households in the LTCI market obtained from Table VI of Hendren (2013).

Given that the implied marginal value of insurance we obtain is sensitive to the assumed curvature parameter, Figure 1.6 plots the range of implied MRS's under each implementation for σ^c ranging from 1 to 4. Focussing first on the weighted sum measure (purple bar), the figure implies that, within this range of commonly assumed relative risk aversion levels, if we believe that individuals are sufficiently risk averse, then *average* marginal benefit from LTCI exceeds the associated barrier imposed by asymmetric information. However, this conclusion needs to be qualified. In particular, comparing the red and blue bars shows that while a higher degree of risk aversion, say $\sigma^c = 3$, is sufficient to imply that the value of providing LTCI to households who are Non-movers exceeds its implicit information cost, the corresponding benefit to households who are Movers does not overcome the same threshold.

Figure 1.6: Sensitivity of implied MRS to assumed curvature parameter value



Note: The figure plots the implied MRS for $1 \leq \sigma^c \leq 4$ under various implementation approaches. The black block corresponds to the naive implied MRS from substituting the average flow consumption response from Column 4 of Table 1.4 into Equation **??**. The red and blue blocks represent the group-specific implied MRS' obtained from substituting the respective flow consumption responses of Non-movers and Movers (Column 8 of Table 1.4) into Equation **??**. The purple block represents the implied MRS from using substituting the respective flow consumption responses of Non-movers and Movers together with the probability of home sale at $j = 0$ into the weighted-sum implementation formula 1.5. The green line and corresponding confidence bands depict the minimum pooled price ratio for rejected households in the LTCI market obtained from Table VI of Hendren (2013).

1.7 Discussion

Accounting for the complementarity effect As shown in Corollary 1.9, to the extent that preferences over food and housing consumption are complementary and under the assumption that second-order complementarity terms are small, our estimates yield an upper bound to the marginal value of insurance. Unfortunately, absent a reliable measure of intensive margin housing responses, at present, we are unable to make a more precise conclusion by implementing the extended implementation formula, Equation 1.9, as part of our main analysis.

Accounting for the state-dependent preferences Corollary 2 shows how we could extend our framework to allow for the case where the marginal utility of consumption differs across health-states. Crucially, the value of θ, the relative preference for consumption when dependent versus health, determines whether our baseline estimates understate ($\theta > 1$) or overstate ($\theta < 1$) the marginal value of insurance. In this respect, existing work using a fully structural approach suggests that the marginal utility of consumption is greater when dependent (De Nardi et al., 2010; Ameriks et al., 2020).[40]

However, the magnitude of the state-dependence term differs quite widely and depends on other simultaneous structural assumptions. Nevertheless, to get an indication of the magnitude of state-dependence implied in this literature, assume no complementarities in preferences over food and housing and suppose that preferences take the separable power form.[41] Then, existing estimates in the literature suggest that θ in Equation 1.12 ranges between 2.45 (given $\sigma^c = 3.81$) (De Nardi et al., 2010) and 8.25 (given $\sigma^c = 5.27$) Ameriks et al. (2020).[42]

Applicability to other insurance contexts While we have framed our theoretical analysis using the terminology of long-term care dependency risk and focussed on housing as a committed consumption good, our conceptual model is sufficiently stylised to be applicable to other insurance settings. As an example, consider the case of unemployment

[40]However, these models typically equate reported expenditure on consumption with consumption. To be precise, we should thus restate these findings as implying that the marginal utility of expenditure on consumption is greater when dependent. For an alternative conceptual definition of the state-dependent preferences and approach to estimating this, see Finkelstein et al. (2013).

[41]This is to match the preference form assumptions used in the literature.

[42]To be precise, these papers broadly assume that utility is given by $\Theta^{1-\sigma}\frac{c^{1-\sigma}}{1-\sigma}$. Thus, θ in our context is given by $\Theta^{1-\sigma}$, which depends on both the magnitude of state-dependence and risk aversion. Furthermore, the value of 8.25 taken from Ameriks et al. (2020) is not an exact mapping as their model allows for a *level* difference in marginal utility across states. By suppressing this, 8.25 understates the full magnitude of dependence implied in the authors' results.

risk and the spousal labour supply. In this context, an "agent" consists of a household with two spouses, c is fully-flexible household consumption and x can be thought of as the secondary earner's leisure consumption (or more precisely, time unemployed). Here, heterogeneity in adjustment costs in spousal labour supply can arise from labour market frictions or from costs to labour market participation (for example through differential access to childcare services).

Given this, the "liquidity channel" implies that the value of insurance against unemployment differs depending on whether or not labour supply of secondary earners are observed to respond to unemployment of primary earners. While this point is obvious, the key departure of the present analysis is to note that a non-adjustment of spouses' labour supply need not equate with the absence of value of consumption smoothing. Rather, some of this inaction may be driven by adjustment costs instead. In this context, our analysis shows that to infer the value of insurance from flexible consumption responses, one needs to appropriately weight the responses of "adjusters" and "non-adjusters" by the likelihood that such adjustments take place.

This short discussion highlights that "housing consumption" in our model essentially plays the role of a self-insurance instrument. In turn, adjustment costs in housing consumption can be interpreted as frictions which impede flexible changes to this self-insurance instrument. When these features are present in a particular context, Lemma 1 and Proposition 1.5 shows how one can continue to apply the consumption-based approach to infer the marginal value of insurance.

1.8 Conclusion

This paper generalises the consumption-based framework for measuring of the value of insurance to account for committed consumption and applies it to study the value of long-term care insurance against nursing home episodes. Our theoretical analysis highlights the need to account for heterogeneity induced by adjustment frictions in committed consumption and provides guidance on the consumption measures relevant for inferring the consumption-smoothing value of insurance. While our work addresses the challenge that committed consumption poses with respect to the consumption-based approach, the resulting framework is still subject to other criticisms of this methodology.[43] As such,

[43] Such as its sensitivity to the assumptions regarding the curvature in preferences and the need to parameterise state-dependence in preferences.

our work can be seen as complementary to the alternative sufficient-statistics approaches proposed in the recent literature (for example Landais and Spinnewijn (2019); Fadlon and Nielsen (2019)). Our empirical application of this approach to the context of LTCI, the first that we are aware of, shows that while households value LTCI significantly above the actuarially-fair rate, not all of them would be willing to pay the markup required to overcome the costs imposed by asymmetric information on insurance providers. In particular, we find that households which make ex-post costly adjustments to housing tend to value insurance less than their non-adjusting counterparts. From a policy perspective, this heterogeneity suggests there may exist efficiency gains from conditioning insurance contracts on ex-post consumption adjustment. However, such forms of conditioning are likely to be infeasible for private insurers and difficult to justify politically for a public insurance scheme.[44] Nevertheless, to the extent that there are observable measures of barrier to consumption adjustment, our findings suggest gains to incorporating these "tagging" measures in LTCI insurance design.

A Proofs of main results

Preliminaries

Let $y_{h,t} = -L$ if $h = B$ and $y_{h,t} = 0$ if $h = G$, using this notation, agent-k's problem given state-vector $\omega_t = (w_t, x_{t-1}, h_0, \ldots, h_t)$ is

$$\begin{aligned} v(\omega_t) = \max_{c,x} \quad & u_h(c,x) - k \times \mathbf{1}(x \neq x_{t-1}) + \beta E_t[v(\omega_{t+1})] && (1.17) \\ s.t. \quad & c + q_t x_t + \frac{w_{t+1}}{r_t} \leq w_t + y_{h,t} + q_t x_{t-1}, && (1.18) \\ and \quad & w_{T+1} \geq \underline{w} \end{aligned}$$

[44]It is likely infeasible to private market actors as it would require monitoring of insurees consumption post contracting.

In addition, given ω_t denote

$$
\begin{aligned}
\left(c^1(\omega_t), x^1(\omega_t)\right) = \arg\max_{c,x} \quad & u_h(c,x) - k + \beta E_t[v(w_{t+1}, x_t, h_0, \ldots, h_{t+1})] \\
s.t. \quad & c + q_t x_t + \frac{w_{t+1}}{r_t} \leq w_t + y_{h,t} + q_t x_{t-1}, \\
& w_{T+1} \geq \underline{w} \\
\left(c^0(\omega_t), x_{t-1}\right) = \arg\max_{c} \quad & u_h(c, x_{t-1}) + \beta E_t[v(w_{t+1}, x_{t-1}, h_0, \ldots, h_{t+1})] \\
s.t. \quad & c + \frac{w_{t+1}}{r_t} \leq w_t + y_{h,t}, \\
& w_{T+1} \geq \underline{w}
\end{aligned}
$$

$\left(c^1(\omega_t), x^1(\omega_t)\right)$ is the optimal consumption bundle given that the agent moves, while $\left(c^0(\omega_t), x_{t-1}\right)$ is the optimal consumption bundle given the agent stays put. These allow us to define the gross utility benefit from moving given ω_t as

$$
\psi(\omega_t) \equiv u_h(c^1(\omega_t), x^1(\omega_t)) - u_h(c^0(\omega_t), x_0) + \beta\Big(E_t[v(w_{t+1}, x_t, h_0, \ldots, h_{t+1})] - E_t[v(w_{t+1}, x_{t-1}, h_0, \ldots, h_{t+1})]\Big)
$$

Thus, given ω_t, agent-k moves at period t if and only if

$$
x_t(\omega_t) \neq x_{t-1} \iff \psi(\omega_t) \geq k\kappa(\omega_t) \tag{1.19}
$$

A.1 Proof of Lemma 1

Let π_B and π_G denote the unconditional probabilities of dependency and good health. Agent-k's ex-ante expected utility can be written as $V(k) = \pi_B v(w_1, x_0, B; k) + \pi_G v(w_1, x_0, G; k)$. Let $G(k)$ be the distribution function of $k \in [0, +\infty)$. We assume that k is independent of ω.

Integrating over k, we obtain the "population" expected utility

$$
\begin{aligned}
\mathcal{V} &= \int V(k) dG(k) \\
&= \int [\pi_B v(w_1, x_0, B; k) + \pi_G v(w_1, x_0, G; k)] dG(k)
\end{aligned}
$$

Next, denote the shorthand $v_h^1(k) = u_h\left(c_h^1, x_h^1\right) - k + \beta v(w_{2,h}, x_h^1, h)$ and $v_h^0(k) = u_h\left(c_h^0, x_0\right) +$

$\beta v(w_{2,h}, x_0, h)$. $v_h^1(k)$ is the indirect utility given agent-k encounters health-h and moves at $t = 1$ and $v_h^0(k)$ is the corresponding utility when they stay-put.

Using this notation, we have

$$\mathcal{V} = \int \Big(\pi_B\Big[\mathbf{1}\{k \le \psi_B\}v_B^1(k) + (1 - \mathbf{1}\{k \le \psi_B\})v_B^0(k)\Big] + \pi_G\Big[\mathbf{1}\{k \le \psi_G\}v_G^1(k) + (1 - \mathbf{1}\{k \le \psi_G\})v_G^0(k)\Big]\Big)\,dG(k)$$

Consider the perturbation (dI, dP). The change in population expected utility from (dI, dP) is

$$\begin{aligned}
d\mathcal{V} =& \int \Big(\pi_B\Big[\mathbf{1}\{k \le \psi_B\}\frac{\partial v_B^1}{\partial w}dI + (1 - \mathbf{1}\{k \le \psi_B\})\frac{\partial v_B^0}{\partial w}dI\Big] \\
&- \pi_G\Big[\mathbf{1}\{k \le \psi_G\}\frac{\partial v_G^1}{\partial w}dP + (1 - \mathbf{1}\{k \le \psi_G\})\frac{\partial v_G^0}{\partial w}dP\Big]\Big)\,dG(k) \\
=& dI \times \pi_B\Big[\int \mathbf{1}\{k \le \psi_B\}dG\frac{\partial v_B^1}{\partial w} + \int (1 - \mathbf{1}\{k \le \psi_B\})dG\frac{\partial v_B^0}{\partial w}\Big] \\
&- dP \times \pi_G\Big[\int \mathbf{1}\{k \le \psi_G\}dG\frac{\partial v_G^1}{\partial w} + \int (1 - \mathbf{1}\{k \le \psi_G\})dG\frac{\partial v_G^0}{\partial w}\Big] \\
=& dI \times \pi_B\Big[\phi_B\frac{\partial v_B^1}{\partial w} + (1 - \phi_B)\frac{\partial v_B^0}{\partial w}\Big] - dP \times \pi_G\Big[\phi_G\frac{\partial v_G^1}{\partial w} + (1 - \phi_G)\frac{\partial v_G^0}{\partial w}\Big]
\end{aligned}$$

The first equality obtains from applying the envelope condition. Note that the perturbation has no extensive margin welfare effect because it only affects the extensive margin decision of agents *on* the margin. And, by optimality of the original allocation, these agents are indifferent between moving and staying-put. The second equality says that *amongst* movers and non-movers of a given h, there is no heterogeneity in the marginal value of wealth. This obtains from our assumption that k is independently distributed, which implies that given that they move, all agents with health-h choose the same consumption bundle. Similarly, all agents with health-h who stay-put choose the same consumption bundle.

At $d\mathcal{V} = 0$, we have

$$\frac{dP}{dI} = \frac{\pi_B}{\pi_G}\frac{\phi_B\frac{\partial v_B^1}{\partial w} + (1 - \phi_B)\frac{\partial v_B^0}{\partial w}}{\phi_G\frac{\partial v_G^1}{\partial w} + (1 - \phi_G)\frac{\partial v_G^0}{\partial w}} \tag{1.20}$$

Define $\mathcal{WTP} = dP/dI$. Then, under the assumption that $\phi_G = 0$, we have

$$\mathcal{WTP} = \frac{\pi_B}{\pi_G}\left[\phi_B \frac{\frac{\partial v_B^1}{\partial w}}{\frac{\partial v_G^0}{\partial w}} + (1-\phi_B)\frac{\frac{\partial v_B^0}{\partial w}}{\frac{\partial v_G^0}{\partial w}}\right] \tag{1.21}$$

Correspondingly, we have

$$\mathcal{MRS} = \phi_B \frac{\frac{\partial v_B^1}{\partial w}}{\frac{\partial v_G^0}{\partial w}} + (1-\phi_B)\frac{\frac{\partial v_B^0}{\partial w}}{\frac{\partial v_G^0}{\partial w}} \tag{1.22}$$

Finally, by optimality of food consumption, the marginal utility of food consumption in each health-state and move-choice is equal to the shadow-value of wealth under that health-state and move-choice. That is,

$$\mathcal{MRS} = \phi_B \frac{\frac{\partial u_B(c_B^1, x_B^1)}{\partial c}}{\frac{\partial u_G(c_G^0, x_0)}{\partial c}} + (1-\phi_B)\frac{\frac{\partial u_B(c_B^0, x_0)}{\partial c}}{\frac{\partial u_G(c_G^0, x_0)}{\partial c}} \tag{1.23}$$

Relation to average of individual MRS

Recall that for each agent, we have

$$MRS(k) = \frac{\frac{\partial v(w_1, x_0, B; k)}{\partial w}}{\frac{\partial (w_1, x_0, G; k)}{\partial w}} \tag{1.24}$$

In general, this can, in turn, be written as

$$MRS(k) = \begin{cases} \frac{\partial v_B^1}{\partial w} \Big/ \frac{\partial v_G^1}{\partial w} & if\ x_B \neq x_0\ and\ x_G \neq x_0 \\ \frac{\partial v_B^1}{\partial w} \Big/ \frac{\partial v_G^0}{\partial w} & if\ x_B \neq x_0\ and\ x_G = x_0 \\ \frac{\partial v_B^0}{\partial w} \Big/ \frac{\partial v_G^1}{\partial w} & if\ x_B = x_0\ and\ x_G \neq x_0 \\ \frac{\partial v_B^0}{\partial w} \Big/ \frac{\partial v_G^0}{\partial w} & if\ x_B = x_0\ and\ x_G = x_0 \end{cases}$$

This implies that integrating over k yields

$$\begin{aligned}\int MRS(k)dG(k) =& \left[\int \mathbf{1}\{k \le \psi_B\}\mathbf{1}\{k \le \psi_G\}dG\right]\frac{\frac{\partial v_B^1}{\partial w}}{\frac{\partial v_G^1}{\partial w}} + \left[\int \mathbf{1}\{k \le \psi_B\}[1-\mathbf{1}\{k \le \psi_G\}]dG\right]\frac{\frac{\partial v_B^1}{\partial w}}{\frac{\partial v_G^0}{\partial w}} \\ &+ \left[\int [1-\mathbf{1}\{k \le \psi_B\}]\mathbf{1}\{k \le \psi_G\}dG\right]\frac{\frac{\partial v_B^0}{\partial w}}{\frac{\partial v_G^1}{\partial w}} + \left[\int [1-\mathbf{1}\{k \le \psi_B\}][1-\mathbf{1}\{k \le \psi_G\}]dG\right]\frac{\frac{\partial v_B^0}{\partial w}}{\frac{\partial v_G^0}{\partial w}} \\ =& \phi_B\phi_G\frac{\frac{\partial v_B^1}{\partial w}}{\frac{\partial v_G^1}{\partial w}} + \phi_B(1-\phi_G)\frac{\frac{\partial v_B^1}{\partial w}}{\frac{\partial v_G^0}{\partial w}} + (1-\phi_B)\phi_G\frac{\frac{\partial v_B^0}{\partial w}}{\frac{\partial v_G^1}{\partial w}} + (1-\phi_B)(1-\phi_G)\frac{\frac{\partial v_B^0}{\partial w}}{\frac{\partial v_G^0}{\partial w}}\end{aligned} \tag{1.25}$$

And under the assumption $\phi_G = 0$,

$$\int MRS(k)dG = \phi_B\frac{\frac{\partial v_B^1}{\partial w}}{\frac{\partial v_G^0}{\partial w}} + (1-\phi_B)\frac{\frac{\partial v_B^0}{\partial w}}{\frac{\partial v_G^0}{\partial w}}$$

Thus, under the assumptions in the main text, $\mathcal{MRS} = \int MRS(k)dG$.

<u>*Case with $\phi_B = 0$ and $\phi_G > 0$*</u>

Suppose that $\phi_B = 0$ and $\phi_G > 0$. Then,

$$\mathcal{MRS} = \frac{\frac{\partial v_B^0}{\partial w}}{\phi_G\frac{\partial v_G^1}{\partial w} + (1-\phi_G)\frac{\partial v_G^0}{\partial w}}$$

And,

$$\int MRS(k)dG = \phi_G\frac{\frac{\partial v_B^0}{\partial w}}{\frac{\partial v_G^1}{\partial w}} + (1-\phi_G)\frac{\frac{\partial v_B^0}{\partial w}}{\frac{\partial v_G^0}{\partial w}}$$

By Jensen's inequality, it is straightforward to see that

$$\int MRS(k)dG > \mathcal{MRS}$$

Intuitively, $\int MRS(k)dG$ weights MRS under choices to move and stay put by their ex-ante probabilities. However, for the case where $\phi_B = 0$ and $\phi_G > 0$, this heterogeneity does not reflect their different capacity for consumption-smoothing against poor health.

Hence taking the average over heterogeneous MRS's overweights the marginal value of providing additional insurance.

A.2 Proofs for results in Section 1.2.3

First, assume no state-dependence. Using the second order Taylor approximations of $\frac{\partial u(c_B^1, x_B^1)}{\partial c}$ around $(c_B^1, x_B^1) = (c_G, x_G)$ and $\frac{\partial u(c_B^0, x_B^0)}{\partial c}$ around $(c_B^0, x_B^0) = (c_G, x_G)$, we obtain.

$$\begin{aligned}
\frac{\frac{\partial u(c_B^1,x_B^1)}{\partial c}}{\frac{\partial u(c_G,x_G)}{\partial c}} \approx & 1 - \left[-c_G \frac{\frac{\partial u^2(c_G,x_G)}{\partial c^2}}{\frac{\partial u(c_G,x_G)}{\partial c}} \right] \frac{c_B^1 - c_G}{c_G} + \left[x_G \frac{\frac{\partial u^2(c_G,x_G)}{\partial x \partial c}}{\frac{\partial u(c_G,x_G)}{\partial c}} \right] \frac{x_B^1 - x_G}{x_G} \\
& + \frac{1}{2} \left[-c_G \frac{\frac{\partial u^2(c_G,x_G)}{\partial c^2}}{\frac{\partial u(c_G,x_G)}{\partial c}} \right] \left[-c_G \frac{\frac{\partial^3 u(c_G,x_G)}{\partial c^3}}{\frac{\partial u^2(c_G,x_G)}{\partial c^2}} \right] \left(\frac{c_B^1 - c_G}{x_G} \right)^2 \\
& + \frac{1}{2} \left[(x_G)^2 \frac{\frac{\partial u^3(c_G,x_G)}{\partial c \partial x^2}}{\frac{\partial u(c_G,x_G)}{\partial c}} \right] \left(\frac{x_B^1 - x_G}{x_G} \right)^2 + \left[c_G x_G \frac{\frac{\partial u^3(c_G,x_G)}{\partial x \partial c^2}}{\frac{\partial u(c_G,x_G)}{\partial c}} \right] \left(\frac{x_B^1 - x_G}{x_G} \right) \left(\frac{c_B^1 - c_G}{c_G} \right) \qquad (1.26)
\end{aligned}$$

$$\frac{\frac{\partial u(c_B^0,x_B^0)}{\partial c}}{\frac{\partial u(c_G,x_G)}{\partial c}} \approx 1 - \left[-c_G \frac{\frac{\partial u^2(c_G,x_G)}{\partial c^2}}{\frac{\partial u(c_G,x_G)}{\partial c}} \right] \frac{c_B^0 - c_G}{c_G} + \frac{1}{2} \left[-c_G \frac{\frac{\partial u^2(c_G,x_G)}{\partial c^2}}{\frac{\partial u(c_G,x_G)}{\partial c}} \right] \left[-c_G \frac{\frac{\partial^3 u(c_G,x_G)}{\partial c^3}}{\frac{\partial u^2(c_G,x_G)}{\partial c^2}} \right] \left(\frac{c_B^0 - c_G}{x_G} \right)^2 \qquad (1.27)$$

Proof of Proposition 1

First suppose $\frac{\partial^2 u(c,x)}{\partial x \partial c} = 0$. Let $\sigma^c \equiv -\tilde{c} \frac{\partial^2 u(\tilde{c},\tilde{x})}{\partial c^2} \Big/ \frac{\partial u(\tilde{c},\tilde{x})}{\partial c}$ and $\gamma^c \equiv -\tilde{c} \frac{\partial^3 u(\tilde{c},\tilde{x})}{\partial c^3} \Big/ \frac{\partial^2 u(\tilde{c},\tilde{x})}{\partial c^2}$ be the elasticities of $\frac{\partial u}{\partial c}$ and $\frac{\partial^2 u}{\partial c^2}$ evaluated at $(\tilde{c}, \tilde{x}) = (c_G, x_G)$. Equivalently, σ^c and γ^c are called the first and second-order measures of the curvature of preferences with respect to c, evaluated at the consumption bundle under good health. Let $\Delta c^0 \equiv c_G - c_B^0$ and $\Delta c^1 \equiv c_G - c_B^1$ denote the change in consumption of non-movers and movers due to dependency.

Substituting Equations 1.26 and 1.27 into Equation 1.25 yields:

$$\mathcal{MRS} \approx 1 + \sigma^c \Big[\phi_B \frac{\Delta c^1}{c_G} + (1 - \phi_B) \frac{\Delta c^0}{c_G} \Big] + \frac{1}{2} \sigma^c \gamma^c \Big[\phi_B \Big(\frac{\Delta c^1}{c_G} \Big)^2 + (1 - \phi_B) \Big(\frac{\Delta c^0}{c_G} \Big)^2 \Big]$$

Proof of Corollary 1

Step 1: Relationship between housing consumption and marginal utility of food consumption

Claim 1. *Let* $\Delta x \equiv x^0 - x^1$, *then*

$$\frac{\partial u(c^1,x^1)}{\partial c} - \frac{\partial u(c^0,x^0)}{\partial c} \lesseqgtr 0 \iff \Delta x \gtreqless 0$$

Proof. From the budget constraint, $c^1 = w_1 + q\Delta x$. By the fundamental theorem of calculus,

$$\frac{\partial u(c^1,x^1)}{\partial c} - \frac{\partial u(c^0,x^0)}{\partial c} = \int_{\chi=0}^{\Delta x} \frac{\partial^2 u(w_1+q\chi, x^0-\chi)}{\partial c^2} q - \frac{\partial^2 u(w_1+q\chi, x^0-\chi)}{\partial x \partial c} d\chi$$

Since $\partial^2 u/\partial c^2 < 0$ and $\partial^2 u \partial x \partial c \geq 0$, we obtain our result. □

Step 2: Relationship between housing and food consumption
Assume $\frac{\partial^2 u(c,x)}{\partial x \partial c} = 0$ and that third and higher order terms in $\frac{\Delta c^1}{c_G}$ and $\frac{\Delta c^0}{c_G}$ are small. Then the second-order approximations of $\frac{\partial u(c_B^1,x_B^1)}{\partial c}/\frac{\partial u(c_G,x_0)}{\partial c}$ and $\frac{\partial u(c_B^0,x_B^0)}{\partial c}/\frac{\partial u(c_G,x_0)}{\partial c}$ are valid. Thus, subtracting Equation 1.27 from Equation 1.26 yields

$$\begin{aligned}\frac{\frac{\partial u(c_B^1,x_B^1)}{\partial c}}{\frac{\partial u(c_G,x_0)}{\partial c}} - \frac{\frac{\partial u(c_B^0,x_B^0)}{\partial c}}{\frac{\partial u(c_G,x_0)}{\partial c}} \approx & \sigma^c\Big[\frac{\Delta c^1}{c_G} - \frac{\Delta c^0}{c_G}\Big] + \frac{1}{2}\sigma^c\gamma^c\Big[\frac{\Delta c^1}{c_G} + \frac{\Delta c^0}{c_G}\Big]\Big[\frac{\Delta c^1}{c_G} - \frac{\Delta c^0}{c_G}\Big] \\ = & \sigma^c\Big[\frac{\Delta c^1}{c_G} - \frac{\Delta c^0}{c_G}\Big]\left(1 + \frac{1}{2}\gamma^c\Big[\frac{\Delta c^1}{c_G} + \frac{\Delta c^0}{c_G}\Big]\right)\end{aligned}$$

Thus,

$$\frac{\frac{\partial u(c_B^1,x_B^1)}{\partial c}}{\frac{\partial u(c_G,x_0)}{\partial c}} - \frac{\frac{\partial u(c_B^0,x_B^0)}{\partial c}}{\frac{\partial u(c_G,x_0)}{\partial c}} \lesseqgtr 0 \iff \frac{\Delta c^1}{c_G} - \frac{\Delta c^0}{c_G} \lesseqgtr 0$$

Next, substituting Claim 1 into the LHS of the relation yields

$$x_G - x_B^1 \gtreqless 0 \iff \frac{\Delta c^1}{c_G} \lesseqgtr \frac{\Delta c^0}{c_G}$$

For consumption responses to negative income shocks, the relation above implies that agents who reduce housing consumption by moving would have a smaller change in food consumption relative to agents who stay put (and leave housing consumption unchanged).

Proof of Corollary 1

Now, allow $\frac{\partial^2 u(c,x)}{\partial x \partial c} \geq 0$. Let $\varepsilon_x^{u_c} \equiv x \frac{\partial^2 u}{\partial x \partial c} \Big/ \frac{\partial u}{\partial c}$ be the elasticity of $\frac{\partial u}{\partial c}$ with respect to x, evaluated at (c_G, x_G). Analogously, let $\varepsilon_x^{u_{cc}} \equiv x \frac{\partial^3 u}{\partial x \partial c^2} \Big/ \frac{\partial^2 u}{\partial c^2}$ and $\varepsilon_c^{u_{xx}} \equiv c \frac{\partial^3 u}{\partial c \partial x^2} \Big/ \frac{\partial^2 u}{\partial x^2}$ be the elasticities of $\frac{\partial^2 u}{\partial c^2}$ with respect to x and $\frac{\partial^2 u}{\partial x^2}$ with respect to c, evaluated at (c_G, x_G). In addition, suppose the agent is consuming at the interior solution to their maximisation problem, then from the FOCs,

$$\frac{\frac{\partial u(c,x)}{\partial x}}{\frac{\partial u(c,x)}{\partial c}} = q$$

Denote $s_G \equiv \frac{q x_G}{c_G + q x_G}$ and $\Delta x^1 \equiv x_G - x_B^1$. Substituting these definitions into Equations 1.26 and 1.27 and using Equation 1.25 yields

$$\begin{aligned} \mathcal{MRS} \approx & 1 + \sigma^c \Big[\phi_B \frac{\Delta c^1}{c_G} + (1-\phi_B)\frac{\Delta c^0}{c_G}\Big] - \varepsilon_x^{u_c}\Big[\phi_B \frac{\Delta x^1}{x_G}\Big] \\ & + \frac{1}{2}\sigma^c \gamma^c \Big[\phi_B \Big(\frac{\Delta c^1}{c_G}\Big)^2 + (1-\phi_B)\Big(\frac{\Delta c^0}{c_G}\Big)^2\Big] \\ & - \frac{1}{2}\sigma_G^x \frac{s_G}{1-s_G}\varepsilon_c^{u_{xx}}\Big[\phi_B\Big(\frac{\Delta x^1}{x_G}\Big)^2\Big] - \sigma^c \varepsilon_x^{u_{cc}}\Big[\phi_B\Big(\frac{\Delta c^1}{c_G^0}\Big)\Big(\frac{\Delta x^1}{x_G}\Big)\Big] \end{aligned} \tag{1.28}$$

Furthermore, rearranging, we obtain

$$\begin{aligned} \mathcal{MRS} \approx & 1 + \sigma^c \Big[\phi_B \frac{\Delta c^1}{c_G} + (1-\phi_B)\frac{\Delta c^0}{c_G}\Big]\left(1 + \frac{1}{2}\gamma^c \frac{\phi_B\Big(\frac{\Delta c^1}{c_G}\Big)^2 + (1-\phi_B)\Big(\frac{\Delta c^0}{c_G}\Big)^2}{\phi_B \frac{\Delta c^1}{c_G} + (1-\phi_B)\frac{\Delta c^0}{c_G}}\right) \\ & - \varepsilon_x^{u_c}\Big[\phi_B \frac{\Delta x^1}{x_G}\Big]\left(1 + \frac{1}{2}\sigma_G^x \frac{s_G}{1-s_G}\frac{\varepsilon_c^{u_{xx}}}{\varepsilon_x^{u_c}}\Big[\frac{\Delta x^1}{x_G}\Big]\right) - \sigma^c \varepsilon_x^{u_{cc}}\Big[\phi_B\Big(\frac{\Delta c^1}{c_G^0}\Big)\Big(\frac{\Delta x^1}{x_G}\Big)\Big] \\ = & 1 + \sigma^c \Big[\phi_B \frac{\Delta c^1}{c_G} + (1-\phi_B)\frac{\Delta c^0}{c_G}\Big]\left(1 + \frac{1}{2}\gamma^c \frac{\phi_B\Big(\frac{\Delta c^1}{c_G}\Big)^2 + (1-\phi_B)\Big(\frac{\Delta c^0}{c_G}\Big)^2}{\phi_B \frac{\Delta c^1}{c_G} + (1-\phi_B)\frac{\Delta c^0}{c_G}}\right) \\ & - \varepsilon_x^{u_c}\Big[\phi_B \frac{\Delta x^1}{x_G}\Big]\left(1 + \frac{1}{2}\sigma_G^x \frac{s_G}{1-s_G}\frac{\varepsilon_c^{u_{xx}}}{\varepsilon_x^{u_c}}\Big[\frac{\Delta x^1}{x_G}\Big] + \sigma^c \frac{\varepsilon_x^{u_{cc}}}{\varepsilon_x^{u_c}}\Big[\frac{\Delta c^1}{c_G^0}\Big]\right) \end{aligned} \tag{1.29}$$

Therefore, if

$$\varepsilon_x^{u_c} + \frac{1}{2}\varepsilon_c^{u_{xx}}\sigma_G^x \frac{s_G}{1-s_G}\frac{\Delta x^1}{x_G} + \varepsilon_x^{u_{cc}}\sigma^c \frac{\Delta c^1}{c_G} \geq 0$$

then the complementarity channel has a negative effect on $\mathcal{MRS}$. This holds for sufficiently small second-order complementarity effects $\varepsilon_x^{u_{cc}}, \varepsilon_c^{u_{xx}}$.

Generalising Corollary 1 to the case with complementarity

For $\frac{\partial^2 u(c,x)}{\partial x \partial c} > 0$, then

$$
\begin{aligned}
&\frac{\frac{\partial u(c_B^1, x_B^1)}{\partial c}}{\frac{\partial u(c_G, x_0)}{\partial c}} - \frac{\frac{\partial u(c_B^0, x_B^0)}{\partial c}}{\frac{\partial u(c_G, x_0)}{\partial c}} < 0 \\
\Rightarrow \sigma^c \Big[\frac{\Delta c^1}{c_G} - \frac{\Delta c^0}{c_G}\Big] &+ \frac{1}{2}\sigma^c \gamma^c \Big[\frac{\Delta c^1}{c_G} + \frac{\Delta c^0}{c_G}\Big]\Big[\frac{\Delta c^1}{c_G} - \frac{\Delta c^0}{c_G}\Big] \\
&< \frac{\Delta x}{x_0}\left[x_0 \frac{\frac{\partial^2 u(c_G, x_0)}{\partial x \partial c}}{\frac{\partial u(c_G, x_0)}{\partial c}}\right]\left(1 - \frac{1}{2}\left[x_0 \frac{\frac{\partial^2 u(c_G, x_0)}{\partial x^2 \partial c}}{\frac{\partial^2 u(c_G, x_0)}{\partial x \partial c}}\right]\frac{\Delta x}{x_0} - \left[c_G \frac{\frac{\partial^2 u(c_G, x_0)}{\partial x \partial c^2}}{\frac{\partial^2 u(c_G, x_0)}{\partial x \partial c}}\right]\frac{\Delta c^1}{c_G}\right)
\end{aligned}
$$

Thus, for

$$
1 - \frac{1}{2}\left[x_0 \frac{\frac{\partial^2 u(c_G, x_0)}{\partial x^2 \partial c}}{\frac{\partial^2 u(c_G, x_0)}{\partial x \partial c}}\right]\frac{\Delta x}{x_0} - \left[c_G \frac{\frac{\partial^2 u(c_G, x_0)}{\partial x \partial c^2}}{\frac{\partial^2 u(c_G, x_0)}{\partial x \partial c}}\right]\frac{\Delta c^1}{c_G} \geq 0
$$

$\Delta c^1 < \Delta c^0$ is sufficient. Otherwise, there exists $\underline{D} < 0$ such that if $\Delta c^1 - \Delta c^0 > \underline{D}$ then

$$
\frac{\frac{\partial u(c_B^1, x_B^1)}{\partial c}}{\frac{\partial u(c_G, x_0)}{\partial c}} - \frac{\frac{\partial u(c_B^0, x_B^0)}{\partial c}}{\frac{\partial u(c_G, x_0)}{\partial c}} \geq 0
$$

Proof of Corollary 2

We now relax the assumption that $\frac{\partial u_B(c,x)}{\partial c} = \frac{\partial u_G(c,x)}{\partial c}$ and $\frac{\partial u_B(c,x)}{\partial x} = \frac{\partial u_G(c,x)}{\partial x}$. However, assume that state-dependence in preferences satisfies

$$
\frac{\frac{\partial u_B(c,x)}{\partial c}}{\frac{\partial u_G(c,x)}{\partial c}} = \frac{\frac{\partial u_B(c,x)}{\partial x}}{\frac{\partial u_G(c,x)}{\partial x}} = \theta, \qquad \text{for each } (c,x) \tag{1.30}
$$

Then

$$
\begin{aligned}
\frac{\partial u_B(c_B^1,x_B^1)}{\partial c} \approx & \frac{\partial u_B(c_G,x_G)}{\partial c}\Bigg\{1-\Bigg[-c_G\frac{\frac{\partial u_B^2(c_G,x_G)}{\partial c^2}}{\frac{\partial u_B(c_G,x_G)}{\partial c}}\Bigg]\frac{c_B^1-c_G}{c_G}+\Bigg[x_G\frac{\frac{\partial u_B^2(c_G,x_G)}{\partial x\partial c}}{\frac{\partial u_B(c_G,x_G)}{\partial c}}\Bigg]\frac{x_B^1-x_G}{x_G}\\
&+\frac{1}{2}\Bigg[-c_G\frac{\frac{\partial u_B^2(c_G,x_G)}{\partial c^2}}{\frac{\partial u_B(c_G,x_G)}{\partial c}}\Bigg]\Bigg[-c_G\frac{\frac{\partial^3 u_B(c_G,x_G)}{\partial c^3}}{\frac{\partial u_B^2(c_G,x_G)}{\partial c^2}}\Bigg]\Bigg(\frac{c_B^1-c_G}{c_G}\Bigg)^2\\
&+\frac{1}{2}\Bigg[(x_G)^2\frac{\frac{\partial u_B^3(c_G,x_G)}{\partial c\partial x^2}}{\frac{\partial u_B(c_G,x_G)}{\partial c}}\Bigg]\Bigg(\frac{x_B^1-x_G}{x_G}\Bigg)^2+\Bigg[c_Gx_G\frac{\frac{\partial u_B^3(c_G,x_G)}{\partial x\partial c^2}}{\frac{\partial u_B(c_G,x_G)}{\partial c}}\Bigg]\Bigg(\frac{x_B^1-x_G}{x_G}\Bigg)\Bigg(\frac{c_B^1-c_G}{c_G}\Bigg)\Bigg\}\\
=&\theta\frac{\partial u_G(c_G,x_G)}{\partial c}\Bigg\{1-\Bigg[-c_G\frac{\frac{\partial u_G^2(c_G,x_G)}{\partial c^2}}{\frac{\partial u_G(c_G,x_G)}{\partial c}}\Bigg]\frac{c_B^1-c_G}{c_G}+\Bigg[x_G\frac{\frac{\partial u_G^2(c_G,x_G)}{\partial x\partial c}}{\frac{\partial u_G(c_G,x_G)}{\partial c}}\Bigg]\frac{x_B^1-x_G}{x_G}\\
&+\frac{1}{2}\Bigg[-c_G\frac{\frac{\partial u_G^2(c_G,x_G)}{\partial c^2}}{\frac{\partial u_G(c_G,x_G)}{\partial c}}\Bigg]\Bigg[-c_G\frac{\frac{\partial^3 u_G(c_G,x_G)}{\partial c^3}}{\frac{\partial u_G^2(c_G,x_G)}{\partial c^2}}\Bigg]\Bigg(\frac{c_B^1-c_G}{c_G}\Bigg)^2\\
&+\frac{1}{2}\Bigg[(x_G)^2\frac{\frac{\partial u_G^3(c_G,x_G)}{\partial c\partial x^2}}{\frac{\partial u_G(c_G,x_G)}{\partial c}}\Bigg]\Bigg(\frac{x_B^1-x_G}{x_G}\Bigg)^2+\Bigg[c_Gx_G\frac{\frac{\partial u_G^3(c_G,x_G)}{\partial x\partial c^2}}{\frac{\partial u_G(c_G,x_G)}{\partial c}}\Bigg]\Bigg(\frac{x_B^1-x_G}{x_G}\Bigg)\Bigg(\frac{c_B^1-c_G}{c_G}\Bigg)\Bigg\}
\end{aligned}
$$

The second equality obtains from the fact that $\frac{\partial u_B(c,x)}{\partial c}=\theta\frac{\partial u_G(c,x)}{\partial c}$ implies

$$
\begin{aligned}
&\frac{\partial^2 u_B(c,x)}{\partial c^2}=\theta\frac{\partial^2 u_G(c,x)}{\partial c^2}, \quad \frac{\partial^3 u_B(c,x)}{\partial c^2}=\theta\frac{\partial^2 u_G(c,x)}{\partial c^3}\\
&\frac{\partial^2 u_B(c,x)}{\partial x\partial c}=\theta\frac{\partial^2 u_G(c,x)}{\partial x\partial c}, \quad \frac{\partial^3 u_B(c,x)}{\partial x\partial c^2}=\theta\frac{\partial^2 u_G(c,x)}{\partial x\partial c^2}, \quad \frac{\partial^3 u_B(c,x)}{\partial c\partial x^2}=\theta\frac{\partial^2 u_G(c,x)}{\partial c\partial x^2}
\end{aligned}
$$

So, using the approximation from 1.26,

$$
\frac{\frac{\partial u_B(c_B^1,x_B^1)}{\partial c}}{\frac{\partial u_G(c_G,x_G)}{\partial c}}\approx\theta\left\{1+\sigma^c\frac{\Delta c^1}{c_G}-\varepsilon_x^{u_c}\Big[\phi_B\frac{\Delta x^1}{x_G}\Big]+\frac{1}{2}\sigma^c\gamma^c\Big(\frac{\Delta c^1}{c_G}\Big)^2-\frac{1}{2}\sigma_G^x\frac{s_G}{1-s_G}\varepsilon_c^{u_{xx}}\phi_B\Big(\frac{\Delta x^1}{x_G}\Big)^2-\sigma^c\varepsilon_x^{u_{cc}}\phi_B\Big(\frac{\Delta c^1}{c_G^0}\Big)\Big(\frac{\Delta x^1}{x_G}\Big)\right.
$$

where the preference curvature measures, $\sigma^c,\gamma^c,\sigma_G^x,\varepsilon_x^{u_c},\varepsilon_x^{u_{cc}},\varepsilon_c^{u_{xx}}$ are evaluated at the consumption bundle under good health (c_G,x_G) and with respect to preferences under good health. Applying the same steps for $\frac{\frac{\partial u_B(c_B^0,x_B^0)}{\partial c}}{\frac{\partial u_G(c_G,x_G)}{\partial c}}$ and substituting into Equation 1.25 yields the expression in the Corollary 2.

A.3 Identifying flow-consumption changes

To identify responses in flow consumption to NH entry, we adapt the approach in **?**. Recall that we obtain information from respondents at specific interviews. Following the notation in the main text, let $j = 0$ denote the interview in which focal NH stay is first reported and let $j = -1$ denote the interview immediately prior.

Generic flow responses to NH entry The HRS contains the dates at which each interview is carried out. Taking the difference between the respective interview dates, we can define the calendar time elapsed between waves as dw. For concreteness, let dw be measured in days (this is without loss). For now, assume interviews are uniformly (across time and respondents) separated by dw.

Let Y_t be defined as the value of the dependent variable *over* the period between $j = t-1$ and $j = t$ (for example, the total consumption over a period of 2×365 days). We are interested in recovering responses in the flow value, y, of the dependent variable.

To do so, first note that at $j = 0$, treated respondents report the number of nights spent in NH between $j = -1$ and $j = 0$. Denoting this number by n, we can define the proportion of time spent in NH between $j = -1$ and $j = 0$ as $\eta = n/dw$.

Under the assumption that Y is linear in the proportion of time spent in NH between interviews, we can write

$$Y_0 = dw[\eta y_B + (1-\eta)y_G]$$

This assumption, together with the assumption that dw is identical for $j = -1$ and $j = 0$ also implies $Y_{-1} = dw \times y_G$. Substituting, we have

$$\begin{aligned} Y_0 &= Y_{-1} + dw \times \eta[y_B - y_G] \\ \frac{Y_0}{dw} &= \frac{Y_{-1}}{dw} + \eta[y_B - y_G] \end{aligned} \tag{1.31}$$

$$\begin{aligned} Y_0 &= Y_{-1} + dw \times \eta[y_B - y_G] + dw \times \eta^2[y_B - y_G] \\ \frac{Y_0}{dw} &= \frac{Y_{-1}}{dw} + \eta[y_B - y_G] + \eta^2[y_B - y_G] \end{aligned} \tag{1.32}$$

We now relax the assumption that dw is uniform and constant across time. This allows us to further align the above identification approach to data in the HRS. With $dw_0 \neq dw_{-1}$, (1.31) reads Assume $Y_{-1} = dw_{-1}y_G$, then

$$\begin{aligned} Y_0 =& dw_0 y_G + dw_0 \times \eta[y_B - y_G] \\ =& \frac{dw_0}{dw_{-1}} \times Y_{-1} + dw_0 \times \eta[y_B - y_G] \\ \frac{Y_0}{dw_0} =& \frac{Y_{-1}}{dw_{-1}} + \eta[y_B - y_G] \end{aligned}$$

Suppose that the researcher observes $\hat{Y}_j$ with a reference period of d days. Then $Y_j = \hat{Y}_j \times (dw_j/d)$. Substituting yields

$$\frac{\hat{Y}_0}{d} = \frac{\hat{Y}_{-1}}{d} + \eta[y_B - y_G]$$

Thus, if the reference period for Y is "since the last interview", then one needs to normalise by the time between interviews.

This implies that $[y_B - y_G]$ can be identified by the coefficient, β, on $n = dw \times \eta$ in the equation

$$\frac{Y_t}{d} = \alpha + \beta\eta + \varepsilon_{it} \tag{1.33}$$

Heterogeneity in flow responses Now consider heterogeneity between movers and non-movers. Following the arguments above, we can write the linear relation between flow and observed values of the dependent variable as

$$\frac{\hat{Y}_0^0}{d} = \frac{\hat{Y}_{-1}^0}{d} + \eta^0[y_B^0 - y_G]$$

Suppose

$$\begin{aligned} \frac{\hat{Y}_{-1}^1}{d} =& \frac{\hat{Y}_{-1}^0}{d} + \gamma \\ \frac{dw_{-1}^1}{d} y_G^1 =& \frac{dw_{-1}^0}{d} y_G^0 + \gamma \end{aligned}$$

Let the flow value for movers be written as $y_B^1 = y_B^0 + \delta$. Then, for movers, we have

$$\frac{\hat{Y}_0^1}{d} = \frac{\hat{Y}_{-1}^1}{d} + \eta^1[y_B^1 - y_G^1]$$

This implies that $[y_B^0 - y_G]$ can be identified by the coefficient β and $[y_B^1 - y_G]$ by $\beta + \delta$ in the equation

$$\frac{Y_{it}}{d} = \alpha + \gamma \times \text{Move}_i + \beta\eta_i + \delta \times \text{Move}_i \times \eta_i + \varepsilon_{it} \tag{1.34}$$

where $\text{Move}_i = 1$ if the respondent is moves in $j = 0$ and $\text{Move}_i = 0$ otherwise and η_i is the proportion of time spent in the NH reported in $j = 0$.

In practice, we modify our baseline estimating equation by estimating

$$\hat{C}_t = \alpha + \delta \times \text{Treat}_i + \sum_{j=-3;j\neq-1}^{3} \gamma_j \times \text{I}_j + \sum_{j=-3;j\neq-1,0}^{3} \beta_j \times \text{I}_j \times \text{Treat}_i + \beta_0 \times \text{I}_0 \times \eta + \zeta X_{it} + \varepsilon_{it} \tag{1.35}$$

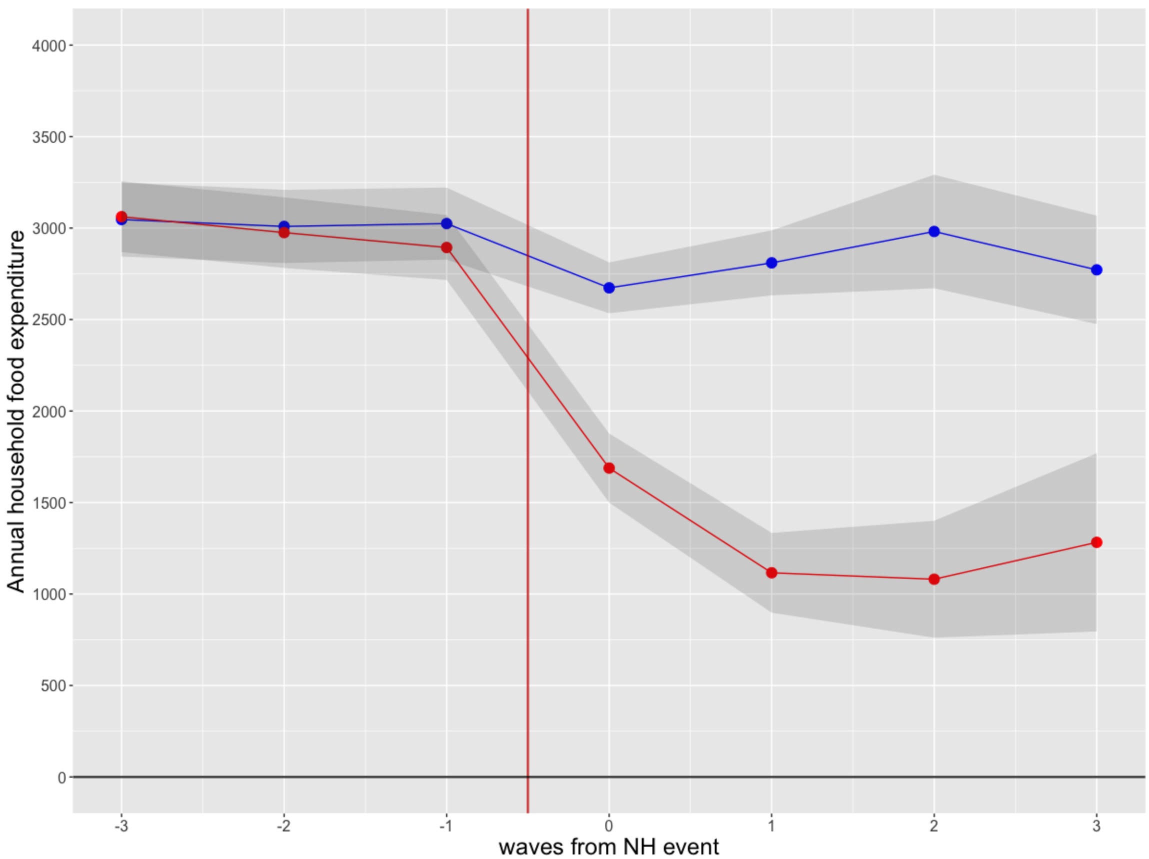
where η is the proportion on time in $j = 0$ spent in the NH.

Alternative Assume that we have observations for treated and control groups.

$$\begin{aligned}
Y_0^{treat} &= dw_0^{treat} y_G + dw_0^{treat} \times \eta[y_B - y_G] \\
&= \frac{dw_0^{treat}}{dw_0^{cont}} \times Y_0^{cont} + dw_0^{treat} \times \eta[y_B - y_G] \\
\frac{Y_0^{treat}}{dw_0^{treat}} - \frac{Y_0^{cont}}{dw_0^{cont}} &= \eta[y_B - y_G] \\
\frac{\hat{Y}_0^{treat}}{d} - \frac{\hat{Y}_0^{cont}}{d} &= \eta[y_B - y_G]
\end{aligned}$$

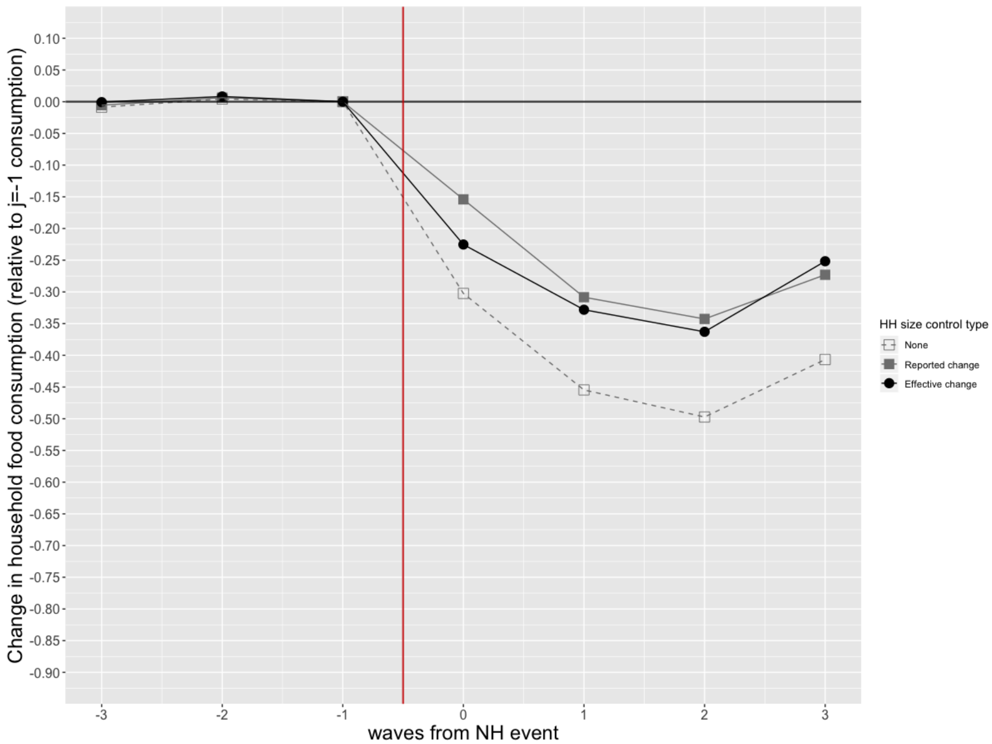
B Additional figures

Figure 1.7: Average household food consumption in treatment and control groups



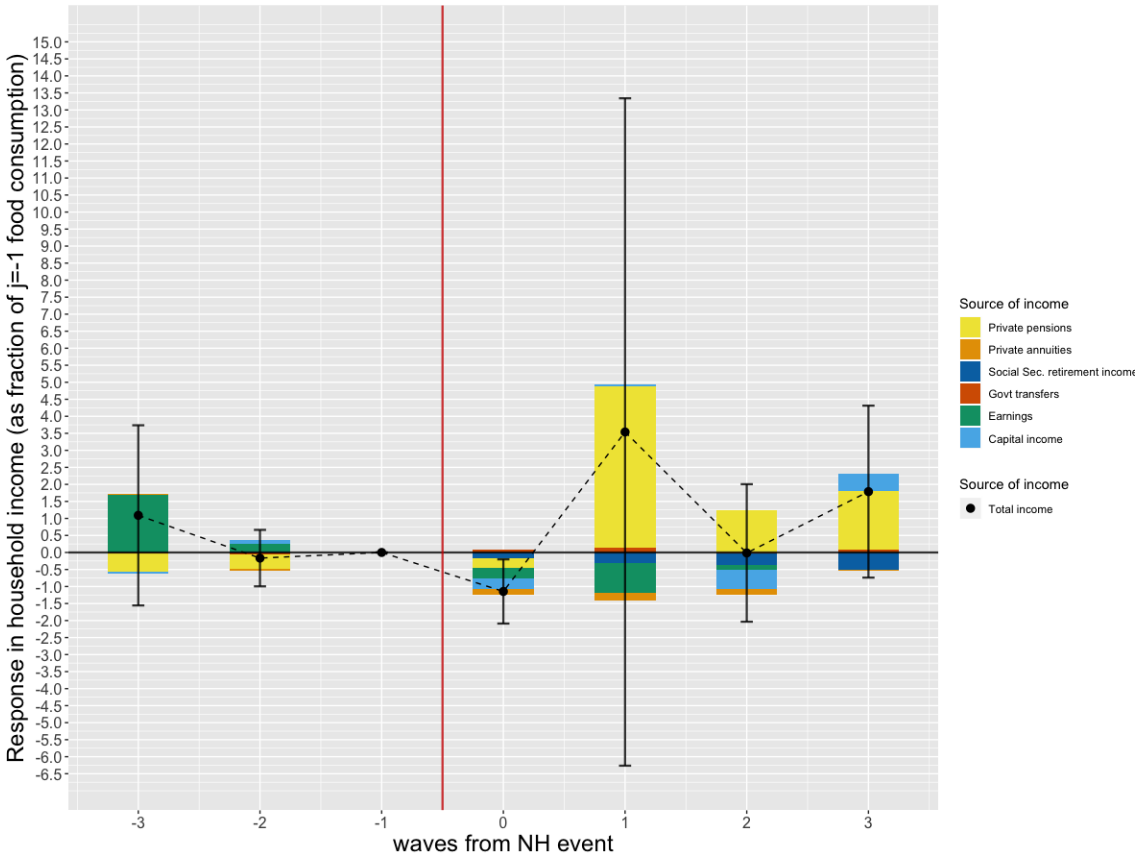
Note: The figure plots the mean annual household food consumption expenditure at event times $j = -3, \ldots, 3$ for treatment (red) and control groups (blue). The grey bands depict 95% confidence intervals for each mean. Food consumption expenditure is defined as the sum of food at home, food outside of home, food delivered to one's home and the value of food stamps received.

Figure 1.8: Consumption response estimates under alternative controls for changes in household size



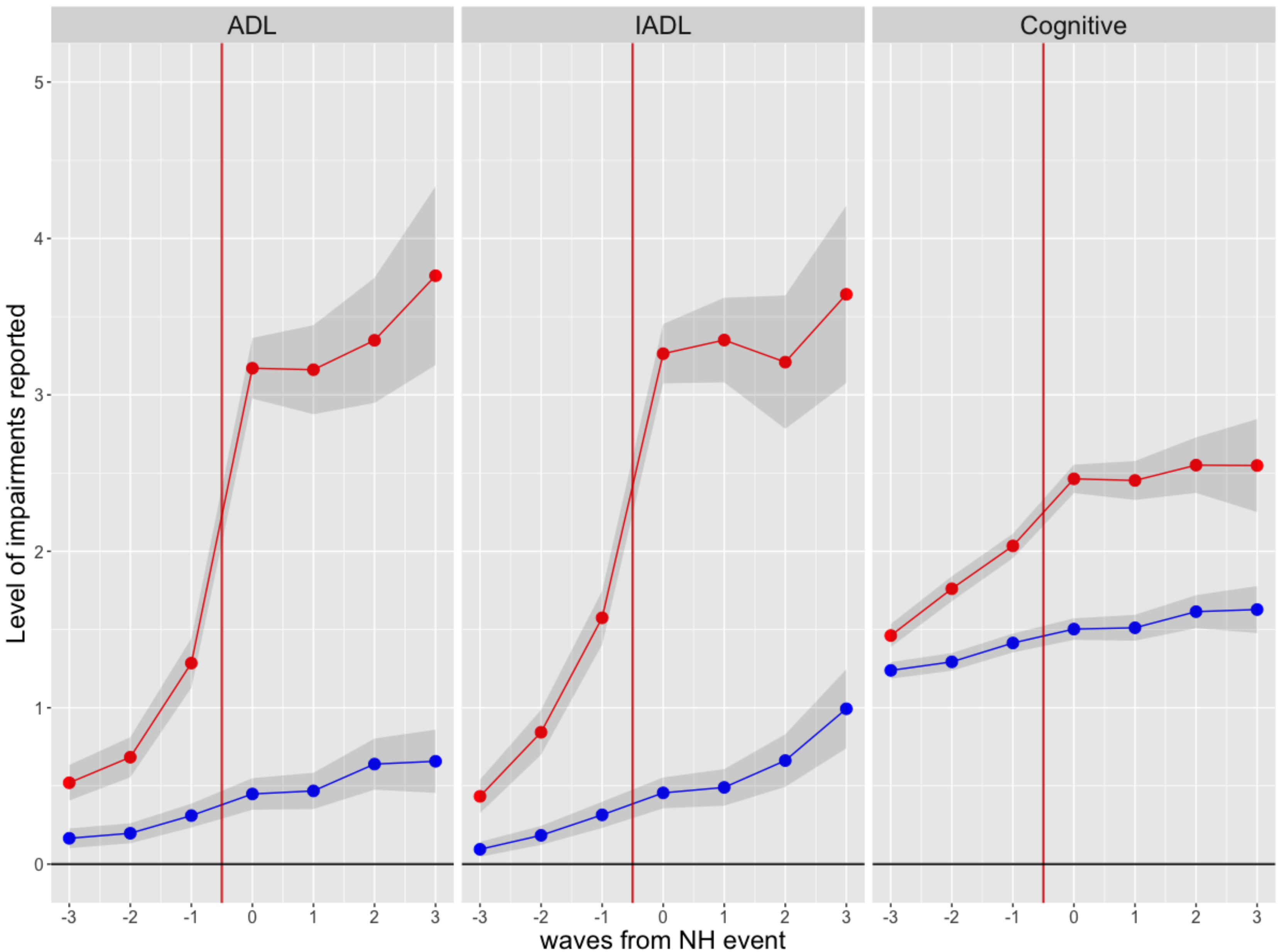
Note: The figure plots the estimates from specification 1.13 with household annual food consumption as the dependent variable. Each point represents β_j, for $j = -3, \ldots, 3$ normalised by average consumption of the treatment group at $j = -1$. All three sets of estimates are conditioned on the full set of demographic and financial covariates given in Table 1.5. Empty squares represent to estimates which do not control for changes in household size changes, solid grey squares represent estimates which control for changes in reported household residents and black circles represent estimates which use the measure of effective change in household size described in the main text.

Figure 1.9: Change in household income by category around NH entry



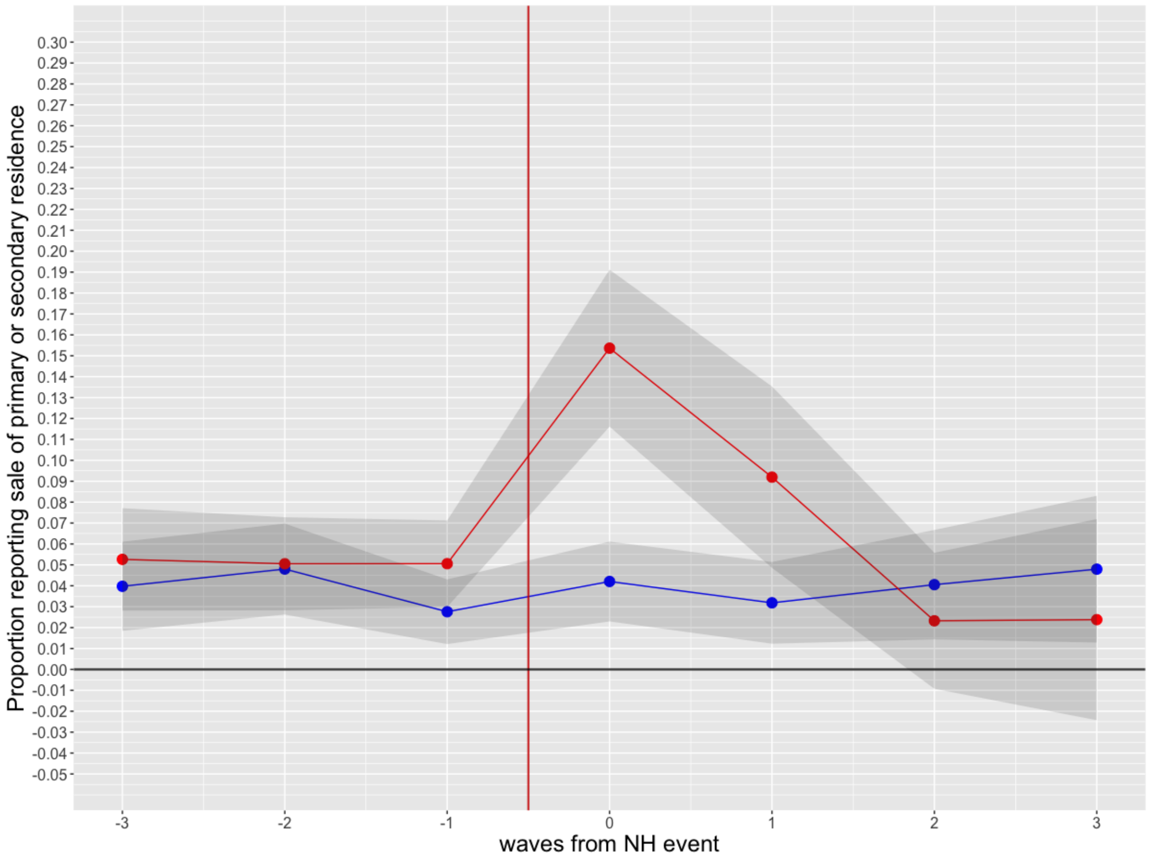
Note: The figure plots the treatment effect estimates from specification 1.13 for various categories of household income for $j = 0, \ldots, 3$. All estimates are expressed as a percentage of average consumption of the treatment group at $j = -1$. 95 % confidence intervals for the change in total income are based on cluster robust standard errors with clustering at the household level.

Figure 1.10: Average level of ADL, IADL and cognitive impairment in treatment and control groups



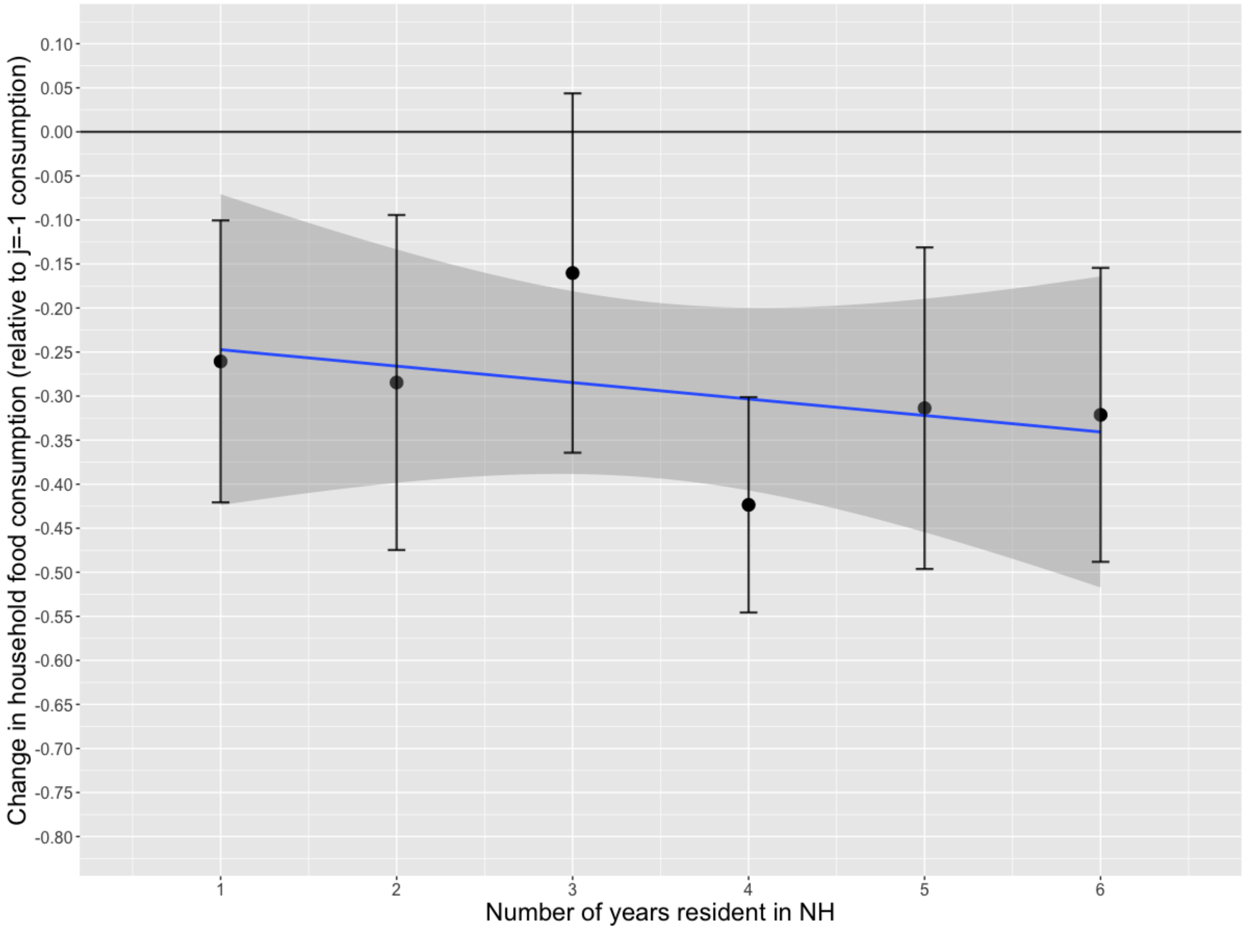
Note: This figure reports the dynamics of physical and cognitive impairments for $j = -3, \ldots, 3$ of respondents for treated (red) and control (groups). Panels A and B plot the average number of reported impa (ranging from 0 to 5) in Activities of Daily Living (ADL) and Instrumental Activities of Daily Living (IADL) for each group. Panel C plots the average level of cognitive impairment, measured on the Lar scale, where 1 corresponds to "No impairment"; 2 to "cognitively-impaired not demented (CIND)" and 3 to "demented".

Figure 1.11: Proportion of households reporting home sales in treatment and control groups



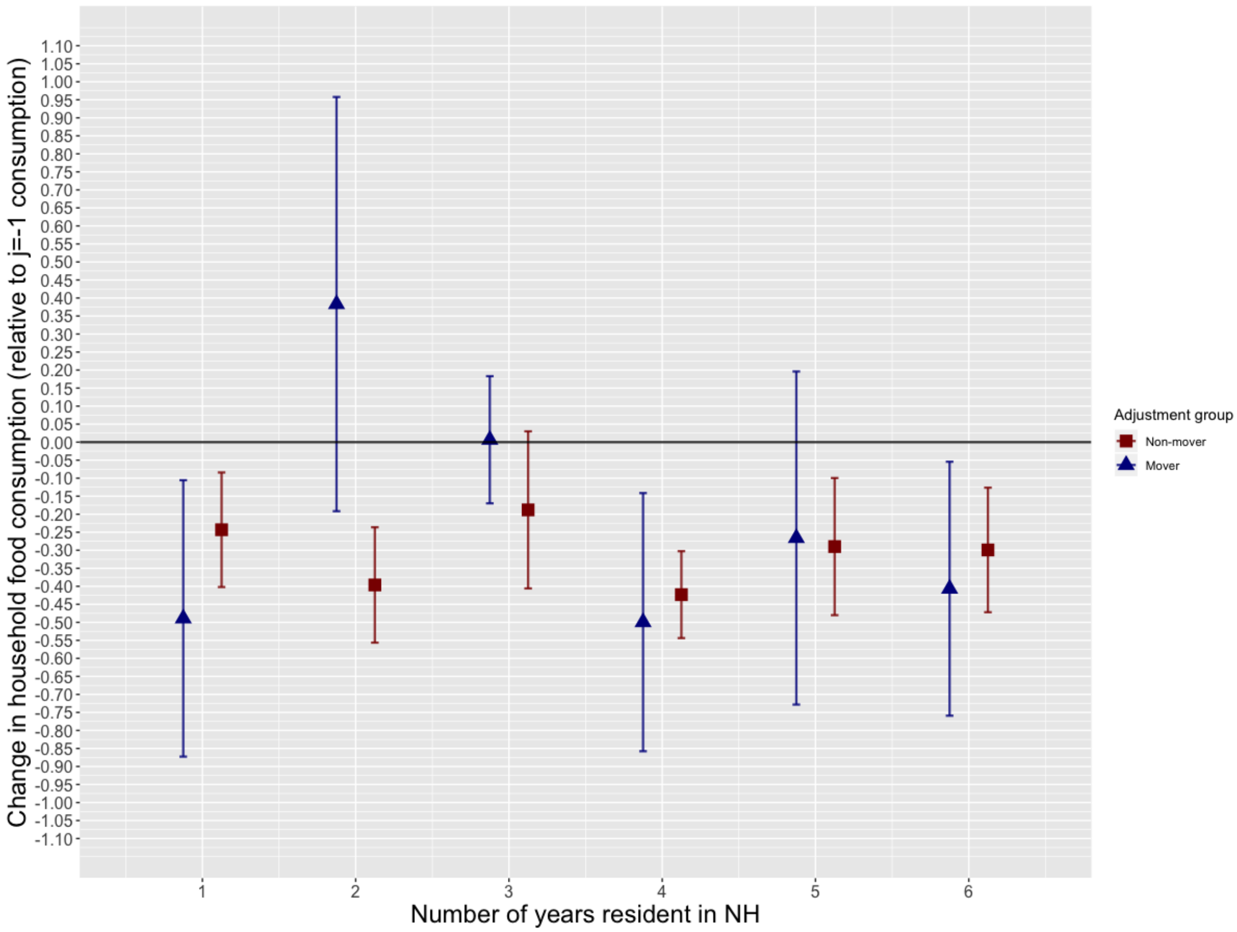
Note: The figure plots the proportion of households reporting a sale of a primary residence or second home at event times $j = -3, \ldots, 3$ for treated (red) and control groups (blue). The grey bands depict 95% confidence intervals for each mean.

Figure 1.12: Non-parametric relationship between reported annual food consumption change and elapsed time in nursing home



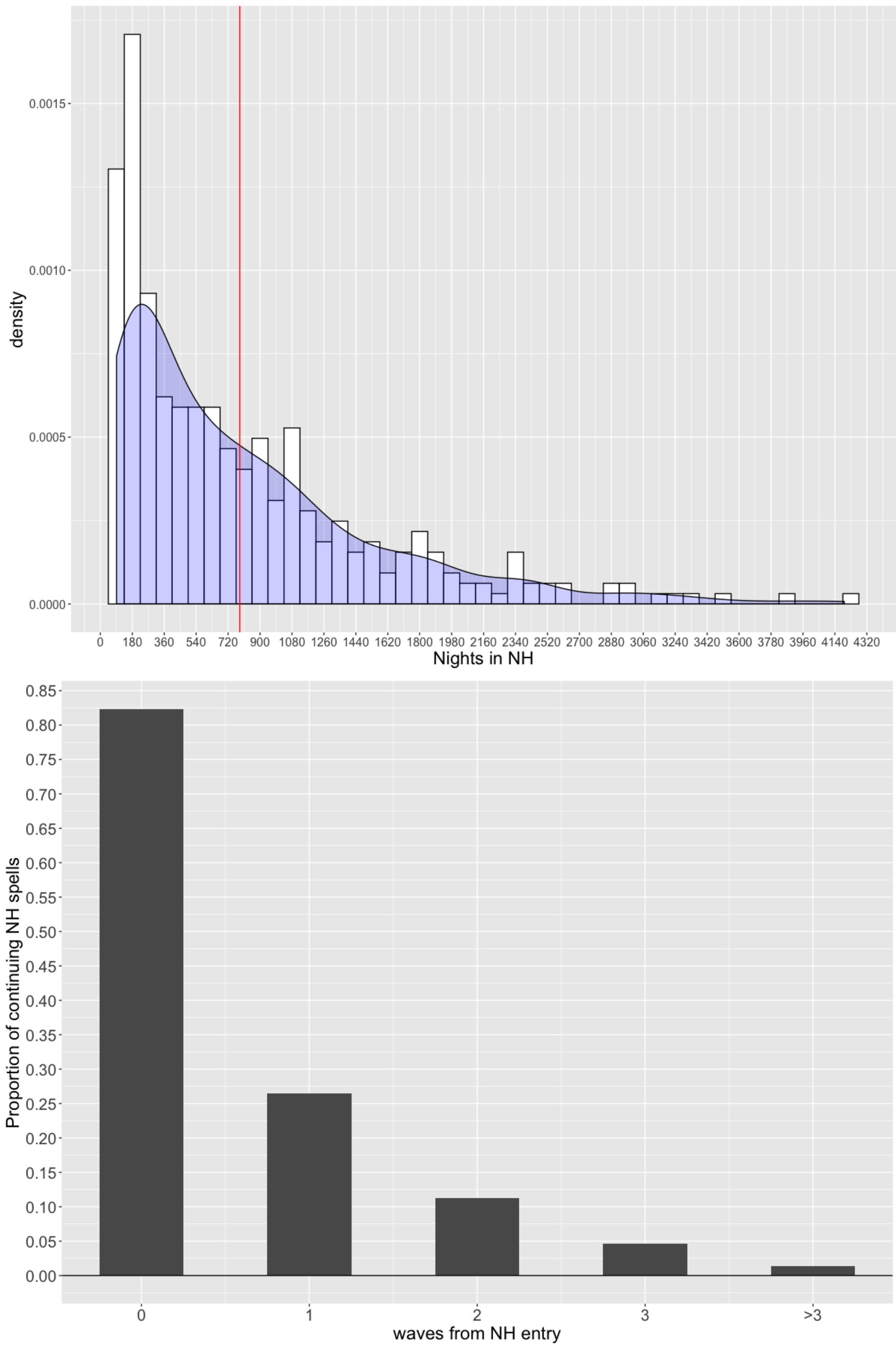
Note: This figure plots the coefficient estimates from Specification 1.15 for the sub-sample of treated households whose respondents are resident in a nursing home at the time of the $j = 0$ interview. These estimates are control for the effective change in household size, as described in the main text, and on the full set of demographic and financial covariates given in Table 1.5. 95 % confidence intervals are based on cluster robust standard errors with clustering at the household level.

Figure 1.13: Non-parametric relationship between reported annual food consumption change and elapsed time in nursing home of “movers” and “non-movers”



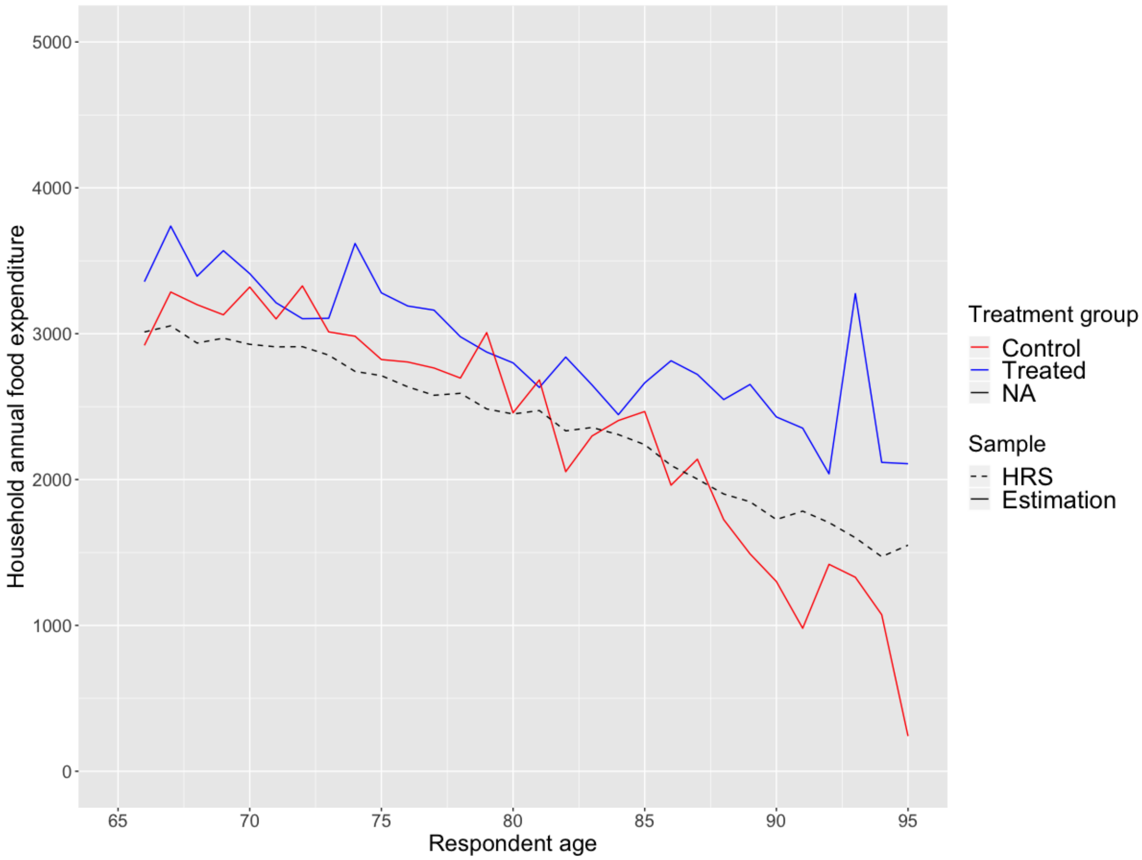
Note: This figure plots adjustment group-specific coefficient estimates from Specification 1.15 for the sub-sample of treated households whose respondents are resident in a nursing home at the time of the $j = 0$ interview. These estimates are control for the effective change in household size, as described in the main text, and on the full set of demographic and financial covariates given in Table 1.5. 95 % confidence intervals are based on cluster robust standard errors with clustering at the household level.

Figure 1.14: Duration of index nursing home spell



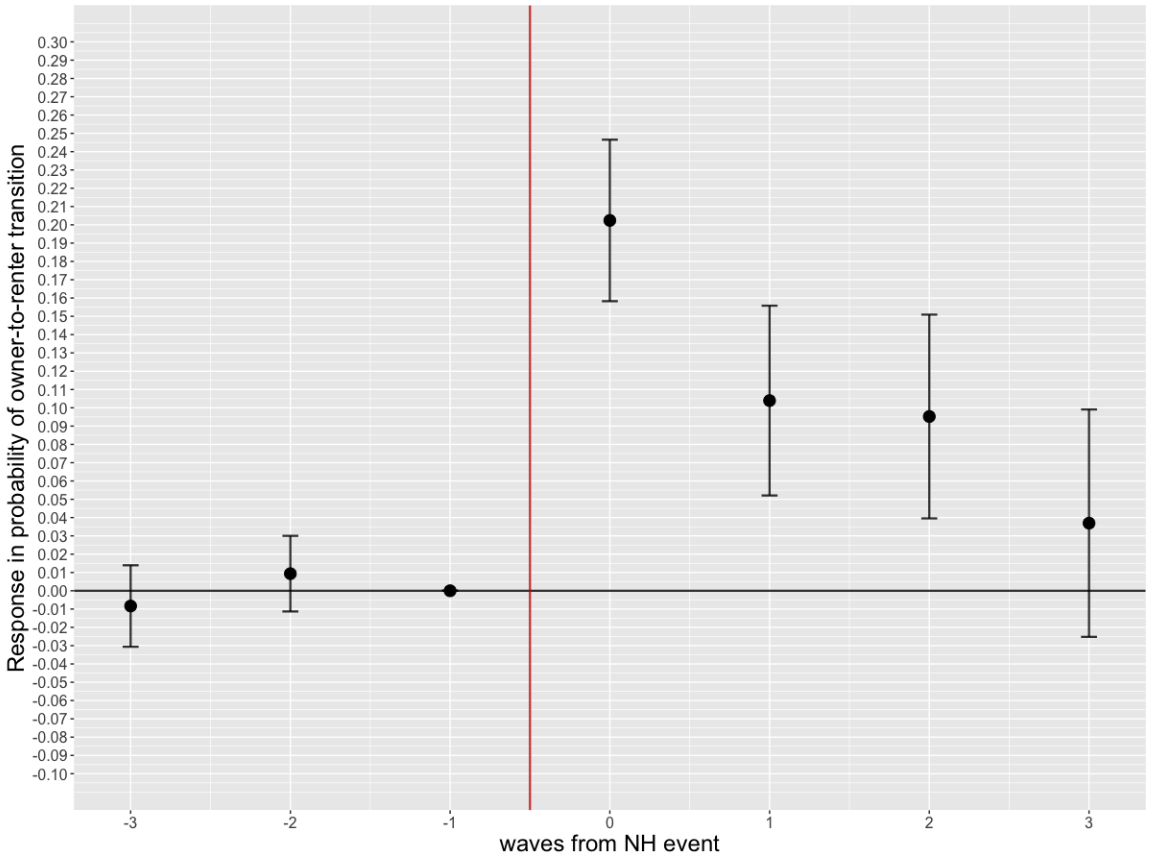
Note: This figure reports the duration of treated respondents' index nursing home spell using two approaches. Panel A plots the histogram and density of the total number of nights spent in nursing home during the index nursing home stay. Each column of the histogram corresponds to a period of 90 nights. Column B plots the proportion of the treated households whose dependent spouse continues to be in the same initial nursing home spell at each event-wave.

Figure 1.15: Age profile of food consumption expenditure in estimation and HRS samples



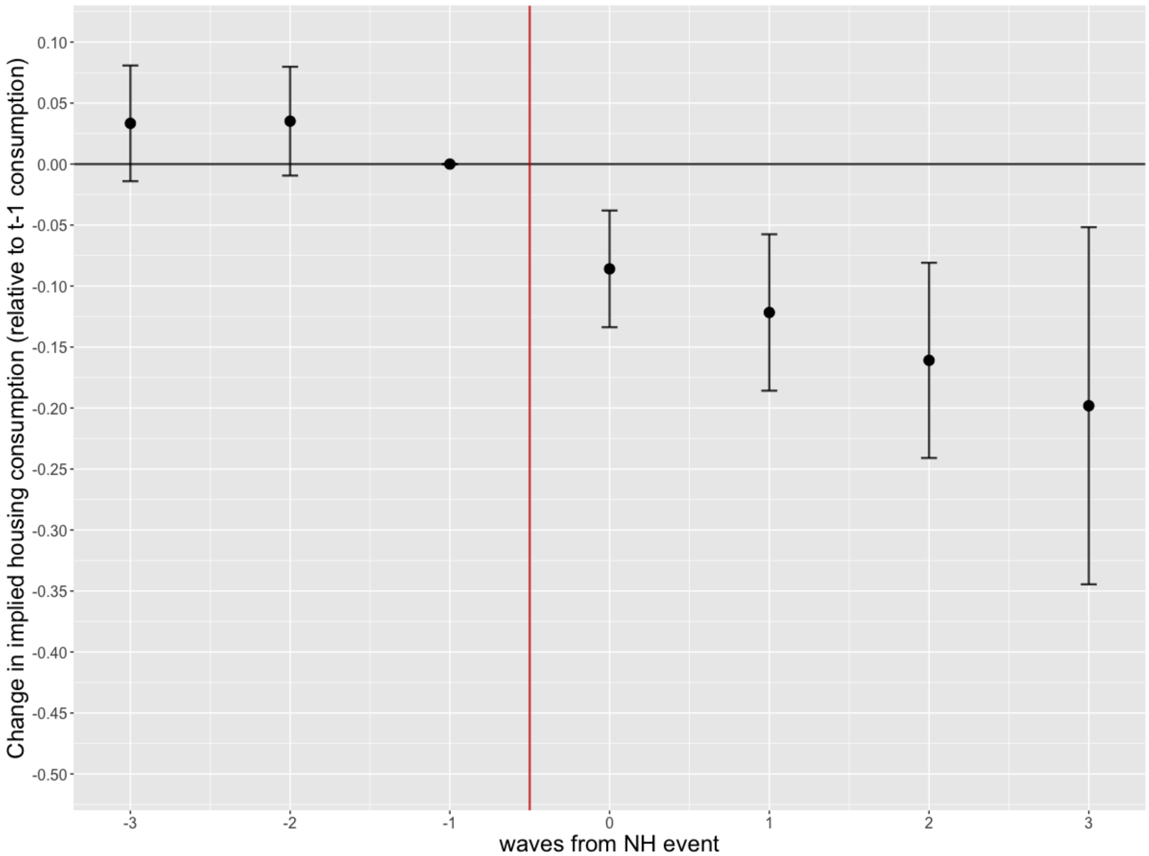
Note: The figure plots the average annual food expenditure against age of respondent in the estimation sample and the main HRS sample.

Figure 1.16: Dynamics of probability home-ownership status change around NH entry



Note: The figure plots the estimates from specification 1.13 with an indicator for a change in home-ownership status as the dependent variable. These estimates are conditioned on the full set of demographic and financial covariates and control for household size changes using Approach 2 (described in the main text). 95 % confidence intervals are based on cluster robust standard errors with clustering at the household level.

Figure 1.17: Dynamics of implied housing consumption around NH entry



Note: The figure plots the estimates from specification 1.13 with implied housing consumption as the dependent variable. Annual housing consumption is taken to be 8 percent of the reported value of one's primary residence. Each point represents β_j, for $j = -3, \ldots, 3$ normalised by average consumption of the treatment group at $j = -1$. These estimates are control for the effective change in household size, as described in the main text, and on the full set of demographic and financial covariates given in Table 1.5. 95 % confidence intervals are based on cluster robust standard errors with clustering at the household level.

C Tables

C.1 Descriptive statistics of analysis sample

Table 1.1: Analysis sample at pre-event wave

	Estimation sample			HRS		
	Mean	Median	St. Dev.	Mean	Median	St. Dev
Respondent age						
Treated	79.025	80	7.736	66.245	65.000	11.409
Control	79.025	80	7.736			
Proportion female						
Treated	0.494			0.555		
Control	0.510					
Proportion white						
Treated	0.862			0.781		
Control	0.874					
Proportion high-sch grad.						
Treated	0.680			0.703		
Control	0.708					
Proportion college grad.						
Treated	0.361			0.403		
Control	0.333					
Proportion partnered						
Treated	0.823			0.615		
Control	0.818					
No. of residents in HH						
Treated	2.322	2	0.853	2.242	2.000	1.256
Control	2.306	2	0.783			
No. of living children						
Treated	3.370	3	2.208	3.201	3.000	2.178
Control	3.356	3	1.978			
Annual food consumption						
Treated	2,894	2,514	1,883	3,034	2,593	2,301
Control	2,991	2,591	2,084			
Total annual income						
Treated	43,358	28,780	53,967			
Control	41,970	32,580	39,611			
Total wealth						
Treated	501,121	221,000	1,204,870	349,332	114,010	842,485
Control	434,662	241,000	660,086			
Home equity share of total wealth						
Treated	0.643	0.606	0.898	0.466	0.375	9.02
Control	0.573	0.606	0.357			
Number of obs.						
Treated		435				
Control		435				

Note: This table presents descriptive statistics of key household characteristics at the pre-event wave ($j = -1$) for treated and control groups in the primary analysis sample of homeowners at $j = -1$. This sample excludes households which at $j = -1$ are enrolled in Medicaid or own private long-term care insurance. The control group is constructed by nearest-neighbour matching on $j = -1$ characteristics. Observations are matched exactly on respondent age and by by propensity score on gender, education level, race, partnered status, census division location, number of household residents, number of living children, total non-housing wealth, value of primary residence, three lags of total household income and total out-of-pocket medical spending and the dependent spouse's actuarial long-term care risk score.

C.2 The effect of nursing home entry on household food consumption

Table 1.2: The effect of nursing home entry on household food consumption

	Annual household food expenditure			
	(1)	(2)	(3)	(4)
Waves from NH entry				
-3	−50.727	−25.138	−16.031	−2.022
	(87.876)	(86.709)	(86.151)	(86.101)
-2	1.118	12.355	18.606	23.881
	(87.469)	(86.617)	(85.450)	(85.664)
0	−1,123.559***	−875.058***	−445.776***	−651.814***
	(119.332)	(107.220)	(115.460)	(107.601)
+1	−1,711.367***	−1,315.478***	−892.262***	−949.382***
	(150.295)	(129.131)	(133.782)	(128.514)
+2	−1,851.569***	−1,438.961***	−991.469***	−1,050.090***
	(216.447)	(185.195)	(182.812)	(191.723)
+3	−1,491.421***	−1,176.764***	−789.830***	−728.160***
	(244.058)	(235.422)	(238.910)	(225.161)
Average cons. at j=-1	2893.78	2893.78	2893.78	2893.78
Survey-wave indicators	✓	✓	✓	✓
Control variables	-	✓	✓	✓
HH-size control: reported change	-	-	✓	-
HH-size control: effective change	-	-	-	✓
Observations	5,718	5,718	5,718	5,718
No. of clusters	870	870	870	870
No. of treated HH	435	435	435	435

Note: *p<0.1; **p<0.05; ***p<0.01
This table reports the fixed-effects regression estimates for the response in household annual food consumption to nursing home entry using Specification 1.13 on our primary analysis sample. Panel 1 displays the estimates for β_j, the coefficient of the interaction between the treatment indicator and event-wave indicators for event-waves $j = -3, \ldots, +3$, taking event-wave $j = -1$ as the reference. Covariates used to control for demographic and financial characteristics are summarised in Table 1.5. Robust standard errors are clustered at the household level.

C.3 The effect of nursing home entry on sale of homes

Table 1.3: The effect of nursing home entry on probability of home sale

	Reported sale of primary or secondary residence			
	(1)	(2)	(3)	(4)
Waves from NH entry				
-3	−0.009	−0.014	−0.012	−0.012
	(0.014)	(0.014)	(0.014)	(0.014)
-2	−0.010	−0.005	−0.005	−0.004
	(0.013)	(0.013)	(0.013)	(0.013)
0	0.097***	0.080***	0.100***	0.088***
	(0.021)	(0.019)	(0.020)	(0.020)
+1	0.053**	0.049**	0.069***	0.058**
	(0.024)	(0.024)	(0.024)	(0.025)
+2	−0.012	−0.018	0.004	−0.007
	(0.021)	(0.022)	(0.023)	(0.023)
+3	−0.010	−0.018	0.005	−0.014
	(0.032)	(0.035)	(0.036)	(0.038)
Average sale prob. at j=-1	0.0506	0.0506	0.0506	0.0506
Survey-wave indicators	✓	✓	✓	✓
Control variables	-	✓	✓	✓
HH-size control: reported change	-	-	✓	-
HH-size control: effective change	-	-	-	✓
Observations	5,718	5,718	5,718	5,718
No. of clusters	870	870	870	870
No. of treated HH	435	435	435	435

Note: *p<0.1; **p<0.05; ***p<0.01
This table reports the fixed-effect regression estimates for the response in the probability of a home sale to nursing home entry using Specification 1.13 on our primary analysis sample. Panel 1 displays the estimates for β_j, the coefficient of the interaction between the treatment indicator and event-wave indicators for event-waves $j = -3, \ldots, +3$, taking event-wave $j = -1$ as the reference. Panel 2 reports the mean consumption at $j = -1$ of treated households which is used in our normalised consumption response plots. Covariates used to control for demographic and financial characteristics are summarised in Table 1.5. Robust standard errors are clustered at the household level.

C.4 Implied flow consumption changes and marginal rate of substitution

Table 1.4: Flow consumption changes and corresponding implied MRS

	Household annual food cons				
	Baseline model				
	(1)	(2)	(3)	(4)	(5)
β_0	$-1{,}709.15^{***}$	$-1{,}288.59^{***}$	-537.88^{**}	-770.06^{***}	$-2{,}113.4$
	(244.29)	(228.02)	(235.39)	(233.15)	(277.40
δ_0					625.61
					(648.31
$\Delta\hat{c}/c_{-1}$	-0.591	-0.445	-0.186	-0.266	
	(0.084)	(0.079)	(0.081)	(0.081)	
$\Delta c^0/c_{-1}$					-0.73
					(0.096
$\Delta c^1/c_{-1}$					-0.514
					(0.206
	Implied MRS (with σ^c				
Naive	1.939	1.644	1.22	1.337	
	(0.184)	(0.149)	(0.112)	(0.123)	
Non-movers					2.264
					(0.236
Movers					1.779
					(0.418
Survey-wave indicators	✓	✓	✓	✓	✓
Control variables	-	✓	✓	✓	-
HH-size control: reported change	-	-	✓	-	-
HH-size control: effective change	-	-	-	✓	-

Note: This table reports the estimated flow change in food consumption in response to nursing home entry. Columns 1 t
the augmented baseline model (Specification 1.14) while Columns 5 to 8 report estimates from the augmented heteroger
1.16). Panel 1 reports the estimate of β_0, the coefficient of the interaction between the indicator for event-wave 0 and th
nursing home reported at $j = 0$ and, for the heterogeneity model, δ_0, the triple interaction between the indicator for event-
time in a nursing home reported at $j = 0$ and the indicator for reported home sale at $j = 0$. Panel 2 reports the correspond
flow consumption with respect to consumption at event-wave -1. Panel 3 reports the implied MRS corresponding to the est
drops assuming that the curvature parameter is $\sigma^c = 1$. This implied MRS is obtained using Equation **??**. Robust standard
the household level are reported in parentheses. Standard errors for implied flow consumption change and MRS are compu

D Additional information

Table 1.5: Summary of covariates used in estimation

Food consumption model	House moving model
Age, $(\text{Age})^2$, $(\text{Age})^3$ Census division location indicator Married/Partnered indicator Total HH wealth, $(\text{Total HH wealth})^2$ Married/Partnered $\times$Total HH wealth Married/Partnered $\times$(Total HH wealth)^2 HH Total OOP medical expenditure	Age, $(\text{Age})^2$, $(\text{Age})^3$ Census division location indicator Married/Partnered indicator Total HH wealth, $(\text{Total HH wealth})^2$ Married/Partnered $\times$Total HH wealth Married/Partnered $\times$(Total HH wealth)^2 HH Total OOP medical expenditure Rental price index at census-region level Survey wave indicator $\times$ Census division indicator

Table 1.6: Summary of variables used in nearest-neighbour matching

Matched by	Variables
Exact	Age at $j = -1$
Propensity score	Gender; Education level (5 levels); Race (3 levels); Married/Partnered status at $j = -1$; Census division location at $j = -1$; No. of HH residents at $j = -1$; No. of living children at $j = -1$ HH total non-housing wealth at $j = -1$; Value of primary residence at $j = -1$ HH total income at $j = -1, -2, -3$; HH total OOP medical expenditure at $j = -1, -2, -3$ Actuarial nursing home risk-score at $j = -1$

Chapter 2

Adverse selection and insurance take-up in a monopoly optimum

2.1 Introduction

This paper proposes and analyses a contract-theoretic monopoly insurance model with unobserved heterogeneity in individuals' risk and preferences for formal insurance. The model extends the classical insurance contracting framework of Stiglitz (1977) by introducing additive heterogeneity in individuals' autarky outside option. Because individuals choose to purchase insurance if and only if the their welfare from being insured exceeds the their welfare in autarky, we interpret this additional dimension of heterogeneity as preferences for participation. These "preferences" could reflect "hassle costs" or "inertia" which are orthogonal to individuals' risk and which affect individuals of the same risk to different extent.[1] Theoretically, introducing heterogeneous participation preferences allows us to study allocations where there is partial take-up within a risk segment. Allowing for such intermediate levels of take-up is important because it enables us to study how the risk-screening motive can affect market size and composition.

To understand our model's implications, recall that in the canonical setting with two risk-types, a monopoly's profit-maximising menu of contracts has the following key features: first, all high-risk individuals obtain the efficient "full insurance" level of coverage. Sec-

[1] For example, two individuals of identical risk may face different time constraints or have different "cognitive bandwidth". To the extent that selecting and purchasing an insurance contract is costly in time or effort, we would expect different take-up behaviour from these individuals.

ond, if they are insured, all low-risk individuals obtain strictly less than full insurance and are left indifferent between autarky and purchasing insurance. Third, whenever the ratio of high to low-risk individuals in the population is sufficiently high, no low-risk individuals are insured (Stiglitz, 1977). More generally, with multiple risks, the classical insurance model implies that the quantity of coverage provided is monotonic in risk with only the highest risks obtaining full insurance. Moreover, when insurance take-up is not complete, non-insurance always begins in the lowest risk groups and take-up behaviour is uniform within a risk group (**?**) Put differently, when comparing insurance levels of a low and high(er) risk group, the canonical model implies that "adverse selection" occurs either along the intensive margin between contracts or along the participation margin, but not both simultaneously.

Our model modifies the above conclusion as follows. First, we find monotonicity in the quantity of coverage is also a feature of any monopoly optimum. Moreover, we also find monotonicity in take-up in our setting. The fact that high-risk individuals are always offered the efficient level of coverage while low-risk individuals never obtain more coverage than high-risks is driven by the risk-screening motive. The risk-screening motive, which affects the monopoly problem via the incentive compatibility constraints also implies that, conditional on being insured, high-risk individuals' consumer surplus is always greater than low-risks' surplus. Given that take-up is determined by the surplus individuals enjoy from purchasing insurance, monotonicity in surplus also implies that take-up rates at the optimum are non-decreasing in risk. That is, for a given incentive compatible allocation, the proportion of high-risk individuals who purchase insurance is weakly greater than the proportion of insured low-risk individuals. Because we assume that participation preferences are orthogonal to risk, our finding that take-up is monotonic in risk is driven entirely by the risk-screening motive.

Second, for each given level of heterogeneity in participation preferences, there is a threshold ratio of high to low-risk individuals above which the monopoly optimum induces partial take-up by low-risks. However, there is in general strictly positive take-up in the low-risk segment. This last finding deviates most starkly from its classical counterpart. Intuitively, the threshold ratio of high to low-risks captures the point at which the profit gain from providing any insurance to *all* low-risk individuals in the population is less than the profit loss from conceding the associated informational rent to all high-risk individuals. When individuals differ in only their risk, a monopoly insurer chooses optimally to offer zero coverage to all low-risk individuals whenever the population distribution exceeds this threshold. However, the ability to induce intermediate levels of take-up in our setting eliminates this kind of all-or-nothing outcome. Rather, the same logic now defines condition under which inducing full take-up is not optimal.

Third, new to our setting, we also find that for a given population distribution of risks, there are levels of preference heterogeneity above which partial take-up by low-risks is optimal. Intuitively, inducing full take-up incurs both a direct cost in the form of surplus conceded to low-risks to induce participation and an indirect incentive cost in the form of information rent conceded to high-risks to maintain risk separation. Because the minimum level of surplus required to induce full-take up is given by the parameter σ, which captures the degree of heterogeneity in participation preferences. An increase in σ increases the cost of inducing full take-up and there is a point at which the cost of inducing take-up by marginal participants exceeds the profit lost if they were left uninsured.

Given the importance of the degree of heterogeneity in participation preferences, we characterise the monopoly optimum over the continuum of possible σ. This analysis shows that a monopoly optimum can be in one of three possible regimes, each with a different configuration of take-up. When the degree of heterogeneity is sufficiently low, the optimum features full take-up by both high and low-risks and low-risks obtain partial insurance. As in the canonical model, low-risks coverage trades off the marginal profit gain from the low-risk segment against the marginal profit loss through informational rent in the high-risk segment. Since the total number of high and low-risks is constant in σ conditional on full take-up, this the coverage provided to low-risks is invariant in σ. Within this regime, low-risks obtain just enough surplus to ensure full take-up while high-risks' surplus is chosen to satisfy the downward incentive compatibility constraint.

At intermediate levels of heterogeneity, the optimum features full take-up by high-risks and partial take-up by low-risks. In this regime, high-risks obtain just enough surplus to induce full take-up in their risk segment. Low-risks' coverage and surplus are jointly chosen to maximise total profit from the low-risks segment subject to satisfying the downward incentive compatibility constraint. The solution to this problem trades off extensive margin gains in profit from a change against intensive margin losses. The extensive margin effect arises because conditional on partial low-risk take-up, changes in surplus offered to this risk segment affects its take-up rate. The intensive margin effect has two components: the change in profit per contract from a change in low-risk surplus holding coverage constant and the change in profit from a change in low-risk coverage holding surplus constant. Finally, we find that coverage provided to low-risks is increasing in σ in this regime. Intuitively, an increase in σ raises the optimal level of high-risk surplus due to the need to maintain full take-up. This, in turn, relaxes the incentive compatibility constraint thereby sustaining higher levels of low-risk coverage at the optimum. When the degree of heterogeneity is sufficiently high, the optimum features partial take-up by both high and low-risks. In this regime, high-risks' surplus, low-risks' surplus and low-risks' coverage are jointly chosen to trade off extensive versus intensive margin effects on profit in each segment. Moreover, low-risk coverage in this case is non-increasing in σ.

2.1.1 Related literature

This paper follows in the tradition of analysing a monopoly insurance problem in a setting with asymmetric information in individuals' risk and where insurers offer menus of coverage-premium contracts Stiglitz (1977); Chade and Schlee (2012). As noted above, the present analysis contributes to this literature by generalising the setting to allow for heterogeneous preferences for becoming insured and characterising the monopoly optimum in this case.

Our introduction of additive heterogeneity in autarky utilities extends the Stligitz-model in the same that Rochet and Stole (2002) extends the classical nonlinear pricing framework of Mussa-Rosen to allow for random participation. Participation within the Rochet-Stole model has also been studied in **?**. The present analysis departs from this literature by considering the case where both individuals' marginal willingness to pay for quality/quantity and the marginal cost of providing this quantity/quality depend on their unobservable "type". The fact that both willingness to pay and cost are type-dependent is characteristic of the insurance problem and sets it apart from the standard product-pricing problem.

Finally, this paper also speaks to the recent literature studying extensive margin patterns of insurance take-up and provision. In this regard, Braun et al. (2019); Chade and Schlee (2020) extend the Stiglitz-model to allow for cost of providing insurance and use this setup to derive optima in which risk groups are denied insurance coverage. The present analysis complements this literature by taking the opposite perspective where individuals face and effective "cost" of becoming insured.

2.2 Model

2.2.1 Timing and risk

There is a unit mass of individuals, each of whom starts with wealth, w, and faces a risk of incurring an exogenous monetary loss, $l > 0$. An insurance contract indemnifies against a fraction, $q \geq 0$, of the loss and costs premium $p \geq 0$. There is a single insurer a menu of coverage-premium contracts. Individuals can either purchase insurance from the two insurers or remain uninsured. We focus on the setting with exclusive contracting and full commitment with the following timing of events: the insurer offers a menu of contracts; individuals observe these offers and choose either to purchase insurance or remain uninsured; losses are realised and payouts are made.

2.2.2 Individuals' preferences

Individuals are heterogeneous and differ along two dimensions denoted by $(\theta, y) \in \Theta \times [0, \sigma]$. In our present analysis, we consider the case with $\Theta = \{\theta_L, \theta_H\}$ where $\theta_L < \theta_H$. f_i is the proportion of θ_i, $i = H, L$, in the population, with $f_H + f_L = 1$. For the sake of exposition, we refer to θ as an individual's risk and y their "participation preferences".

Risk preferences Let $U(q, p; \theta)$ denote the expected utility of an individual with risk θ under contract (q, p). Let $v(q;\theta)$ be defined implicitly by $U(q, v(q;\theta); \theta) = U(0, 0; \theta)$. Conditional on purchasing insurance, $v(q;\theta)$ is the maximum amount that a θ-risk individual is willing to pay for q-level of coverage and is a money-metric representation of an individual's risk preferences. Note that by definition, $v(0;\theta) = 0$. We call $v(q;\theta)$ the "willingness to pay" for q by θ-risks and hereafter work with $v(q;\theta)$ instead of U in our analysis. Assume risk preferences are such that $v(q;\theta)$ has the following properties.

Assumption 1. *Assume $v(q;\theta)$ is twice continuously differentiable and satisfies*

1. $v_q(q;\theta) \geq 0$, $v_{qq}(q;\theta) < 0$ *(concavity)*
2. $v_\theta(q;\theta) \geq 0$ *(risk-dependence)*
3. $v_{q\theta}(q;\theta) > 0$ *(single-crossing in (q, θ))*

Assumption 1 requires individuals' willingness to pay to be increasing and strictly concave in q (Point 1). Furthermore, it assumes that willingness to pay for each *level* of coverage is non-decreasing in individuals' risk (Point 2) and that the *marginal* willingness to pay for each additional unit of coverage is increasing in risk (Point 3). Below, we show that Assumption 1 is satisfied in the most commonly analysed insurance environments.

Preferences for participation Let $u_i \equiv v(q_i;\theta_i) - p_i$ denote the gross surplus enjoyed by a θ_i-risk individual from (q_i, p_i). We assume that participation preferences enter additively in wealth-space. Specifically, assume that a θ_i-individual prefers purchasing (q_i, p_i) over autarky if and only if

$$v(q_i;\theta_i) - p_i - y \geq v(0;\theta_i) \iff u_i \geq y$$

This condition says that an individual chooses to purchase insurance if and only if their surplus from participation is greater than y. Note that y is money-metric like p and $v(q;\theta)$ and does not directly depend on the contract itself. We will focus on the case where $y \geq 0$. Intuitively, y captures the cost, over and above premiums, of being formally insured. Such costs could arise from actual financial costs (such as sign-up fees, or medical certification costs). Alternatively, y could also capture non-pecuniary "hassle" or search costs. y is therefore a reduced form measure of these costs, denominated in wealth.

We assume that y is independent of θ and has a distribution with bounded support $[0,\sigma]$[2]. For our analysis, we focus on the normalised variable, $z \equiv \frac{y}{\sigma}$, which has support $[0,1]$. Let $G(z)$ denote the distribution function of z. Assume that $G(z)$ is log-concave. For a given insurance contract which offers u, let $G(u/\sigma)$ is the probability that a given individual in the population purchases this contract. We hereafter refer to $G(u/\sigma)$ as the take-up rate of a contract offering u. Given this interpretation, we can define the elasticity of take-up as

$$\eta(u;\sigma) = \frac{u}{\sigma}\frac{G'(u/\sigma)}{G(u/\sigma)}$$

2.2.3 Insurer's profit

Insurers' profit depends only on the contract offered and the insured's risk. Let $c(q;\theta)$ be the cost of providing coverage q to an individual of risk θ.

[2]If y were distributed with unbounded support then, by construction, some individuals would always prefer being uninsured. A bounded support ensures that all individuals would choose to be insured if offered a sufficiently high but finite surplus.

Assumption 2. *Assume $c(q;\theta)$ is twice continuously differentiable and satisfies*

1. $c(0;\theta) = 0$
2. $c_q(q;\theta) > 0$, $c_{qq}(q;\theta) \geq 0$ *(convexity)*
3. $c_\theta(q;\theta) > 0$ *for* $q > 0$ *(risk-dependence)*
4. $c_{q\theta}(q;\theta) > 0$

Point 1 assumes zero costs when no coverage is provided while Point 2 assumes that costs are strictly increasing and convex in coverage. The third and fourth points imply that the total (Point 3) and marginal cost (Point 4) of providing insurance is increasing in risk. Points 3 and 4 distinguish our setting from the standard nonlinear product pricing problem (in the vein of Mussa-Rosen) where costs are independent of consumers' "type".

Surplus from insurance We denote the surplus from providing insurance coverage, q, to an individual with θ by $s(q;\theta) \equiv v(q;\theta) - c(q;\theta)$. The insurer's profit from providing (q_i, p_i) to a θ_i-individual is $s(q_i;\theta_i) - u_i$. Intuitively, *conditional* on contracting, the surplus from insurance is split between the insurer and the insured. Note that our definition of $s(q;\theta)$ does not account for costs of participation embodied by y. This definition is appropriate as we are considering a monopoly in the private market, whose profit does not depend directly on y.

Because of the concavity assumptions in Assumptions 1 and 2, we know that $s(q;\theta)$ is strictly concave and hence has a unique maximiser $q_i^* = \arg\max_q s(q;\theta_i)$. For brevity, we let s_i^* denote the maximised surplus for each i. In addition to Assumptions 1 and 2, we further assume that $s(q;\theta)$ satisfies the following property.

Assumption 3. $q_i^* = q^* > 0$ *for each* $\theta_i \in \Theta$.

Assumption 3 says that the surplus-maximising level of coverage exists, is positive and does not depend on individuals' risk. It is satisfied in standard models of insurance contracting with unobservable risk. However, this assumption may be violated in models where risk aversion is correlated with risk and in other behavioural models with non-standard preferences.

To demonstrate how Assumptions 1, 2 and 3 relate to the existing literature, the following two examples show that these assumptions hold in the most commonly analysed insurance environments in the literature.[3]

Example 1: Binary risk with decreasing absolute risk aversion (DARA) Suppose individuals face a given loss of $l > 0$ with probability $\pi(\theta)$ and this probability is strictly increasing in θ. That is, $\pi(\theta_H) > \pi(\theta_L)$. Assume that individuals' Bernoulli utility is increasing, concave and satisfies decreasing absolute risk aversion (DARA). Then Assumptions 1, 2 and 3 are satisfied.

Example 2: Normally-distributed loss with constant absolute risk aversion (CARA) Suppose individuals face a Normally-distributed loss of $l \sim \mathcal{N}(\theta, var(l))$ so the expected loss is increasing in the risk parameter, θ. Assume that individuals' Bernoulli utility satisfies constant absolute risk aversion (CARA). Then Assumptions 1, 2 and 3 are satisfied.

2.3 Monopoly insurance with preferences for participation

2.3.1 Observable risk benchmark

We start by considering the case where θ_i is observable but y is not. Under this "observable risk benchmark", the monopoly solves $\max_{u_i} f_i G(u_i;\sigma)[s_i^* - u_i]$ separately for each θ_i. The following result states that a monopoly optimum exists for all $\sigma \geq 0$ and characterises the optimum for all such parameter values.

Lemma 3. *Suppose θ is observable but y is not. Assume $G(u/\sigma)$ is log-concave and that Assumptions 1, 2 and 3 hold. Then there exists a solution to the monopoly problem for all $\sigma \geq 0$. Furthermore, there exists a unique threshold value, $\hat{\sigma}_i \in [0, s_i^*]$, such that*

1. *For $\sigma \in [0, \hat{\sigma}_i]$, the monopoly optimum is given by $u_i = \sigma$.*
2. *For $\sigma > \hat{\sigma}_i$, the monopoly optimum is given by $u_i = \frac{\eta(u_i;\sigma)}{1+\eta(u_i;\sigma)} s_i^*$.*

Moreover, for $\sigma > \bar{\sigma}_i$, the take-up rate, $G(u_i/\sigma)$, is strictly decreasing in σ and insureds' gross surplus, u_i, is increasing (decreasing) in σ if and only if $\eta(u;\sigma)$ is decreasing (increasing) in u.

[3] Detailed proofs for both examples provided in Supplementary Appendix.

With heterogeneous preferences for participation a monopoly optimum can be in one of two regimes. Moreover, there is a unique threshold value of $\sigma \geq 0$ which determines which of the above regimes is optimal.

When the degree of heterogeneity is sufficiently low, a monopolist optimally induces full take-up by offering all individuals just enough surplus to keep those with the weakest participation preferences indifferent between autarky and taking up insurance. When the degree of heterogeneity is sufficiently high, the monopoly optimally induces partial take-up. In this regime, contracts are priced to trade off intensive-margin losses in profit per contract against the extensive-margin gains in take-up. The surplus from trade, s_i^*, is split between the monopolist and the insured, with the latter's surplus increasing in the elasticity of take-up. Log-concavity of G ensures there exists a unique solution to the monopoly insurance problem for all $\sigma \geq 0$ in this benchmark setting.

The intuition behind the threshold condition for optimality between regimes is as follows. When $\sigma \geq 0$ is small, the amount of surplus conceded to individuals in order to induce full take-up is small. However, as σ rises, there is a point beyond which the (intensive margin) profit obtained from insuring individuals along the participation margin is less than the profit gained from pricing these marginal individuals out but increasing the profit margin from each remaining participant. Put another way, the negative market size effect of lowering u_i is outweighed by its positive effect on profit per contract.

The final point in Lemma 3 says that take-up is non-increasing in the heterogeneity parameter, σ, for all $\sigma \geq 0$. Intuitively, an increase in heterogeneity in participation preferences makes it more expensive for the insurer, in terms of surplus forgone, to induce the same level of take-up. In contrast, the relationship between insured's surplus, u_i, and σ can go in either direction and depends crucially on the shape of $G(z)$.

2.3.2 Monopoly insurance problem with asymmetric information

We now consider the full asymmetric information setting where (θ_i, y) are unobservable to the insurer, but f_H, f_L and $G(z)$ are known. The monopoly insurance problem is

$$\begin{aligned}
\max_{\{(u_i,q_i)\}_{i=H,l}} \quad & f_H G(u_H/\sigma)[s(q_H;\theta_H) - u_H] + f_L G(u_L/\sigma)[s(q_L;\theta_L) - u_L] && (\mathcal{P}^m) \\
s.t. \quad & u_H \geq u_L + \Delta(q_L) && (IC_H) \\
& u_L \geq u_H - \Delta(q_H) && (IC_L) \\
& u_L \geq 0,\ q_L \geq 0
\end{aligned}$$

A solution to the canonical monopoly insurance model with only heterogeneity in risk yields has three key features. First, all high-risk individuals always obtain the efficient level of insurance coverage. Second, low-risk individuals always obtain less than the efficient level of coverage and enjoy a surplus equal to their autarky utility. Third, whenever the population distribution of high-risks exceeds a threshold level low-risk individuals obtain zero insurance coverage while high-risk individuals enjoy a surplus equal to their autarky utility. Our first set of results highlights how the above implications change when we introduce heterogeneity in preferences for participation.

Proposition 2. *Suppose a solution to the monopoly insurance problem exists and let it be denoted by $C^m \equiv \{(u_H, q_H), (u_L, q_L)\}$.*

- *For all $\sigma \geq 0$, $q_H \geq q_L$, $u_H \geq u_L$ and $q_H = q^*$. This implies that take-up rates are non-decreasing in risk. i.e. $G(u_H/\sigma) \geq G(u_L/\sigma)$.*
- *For all $\sigma \geq 0$, $u_L \leq \sigma$.*
- *For each $\sigma \geq 0$, there exists a threshold proportion of high-to-low risks, $\bar{\phi} > 0$, such that $u_H \leq \sigma$ if and only if $\frac{f_H}{f_L} \geq \bar{\phi}$.*
- *For each $\frac{f_H}{f_L} \geq 0$, there exists a threshold level of heterogeneity, $\bar{\sigma}_1 \geq 0$, such that $u_H > \sigma$ if and only if $\sigma < \bar{\sigma}_1$. Moreover there exists a threshold value, $\bar{\phi}$, such that for all $\frac{f_H}{f_L} \geq \bar{\phi}$, we have $\bar{\sigma}_1 = 0$.*

Point 1 of Proposition 2 says that high-risk individuals are always offered the efficient level of coverage while low-risk individuals obtain weakly less coverage than high-risks. Furthermore, high-risks enjoy greater surplus from their offered insurance contract than low-risks do from theirs. Monotonicity of q_i and u_i is a consequence of the incentive compatibility constraints which, in turn, stem from the hidden risk problem and the insurer's risk-screening motive. In addition to these between-contract patterns, monotonicity in u_i in our context also implies that take-up is weakly increasing in risk. The fact that $q_H \geq q_L$ and $G(u_H/\sigma) \geq G(u_L/\sigma)$ means a monopoly optimum could simultaneously feature "adverse" selection between contracts and in take-up. Also, unlike the canonical monopoly setting, there need not be full insurance take-up by high-risks.

Point 2 of Proposition 2 says that low-risk individuals enjoy a surplus no greater than σ when they purchase insurance. Because σ is the minimum participant surplus required to induce full take-up, $u_L \leq \sigma$ implies that a monopolist never concedes more surplus to low-risk individuals than is required to induce full take-up by this group. This is a generalisation of the classic result that low-risks obtain their autarky reservation utility at a monopoly optimum in a Stiglitz-model.

Point 3 modifies the third classic monopoly result mentioned above the same way Point 2 changes the second – the risk distributiion threshold condition now defines cases where there is (weakly) partial take-up by high-risk individuals. A consequence of this generalisation is that having a population risk distribution which exceeds the threshold, $\bar{\phi} > 0$, does not necessarily imply that low-risks are "uninsured", as in the canonical case. Intuitively, if high-risk individuals are sufficiently prevalent in the Stiglitz-monopoly setting, then the total profit forgone from conceding informational rent to the entire population of high-risks outweighs the total profit gained from providing any insurance to low-risks. Therefore, a monopolist optimally offers $q_L = 0$ and $u_L = 0$, so there is zero insurance provision to low-risks. Also, $q_H = q^*$ and $u_H = 0$ so high-risks obtain only their autarky utility. However, with heterogeneous participation preferences, an insurer is able to induce partial but strictly positive take-up within a group with identical risk. This means the insurer no longer faces a trade-off between providing coverage to all low-risk individuals and conceding rent to all high-risk individuals. Instead, it can choose to induce partial take-up by high-risk individuals, thereby reducing the total information rent conceded.

Point 4 of Proposition 1 mirrors Point 3 and says that for any given distribution of risks in the population, there is a threshold level of heterogeneity, $\bar{\sigma}_1 \geq 0$ above which high-risks enjoy no more than σ level of surplus when they are insured and there is partial take-up amongst low-risk individuals. The logic of this result derives from the trade-off between intensive and extensive margin profit gains. Intuitively, higher levels of σ mean that more surplus has to be offered to low-risks to induce full take-up. By the incentive compatibility condition for high-risks, this also means more information rent must be conceded to high-risks. Together, these forces imply that for any given distribution of risks in the population, there is a level of heterogeneity above which a monopolist optimally induces partial take-up by low-risks and offers high-risks at most participant surplus equal to σ.

2.3.3 Characterisation

The previous section discussed the properties that any solution to the monopoly insurance problem must satisfy. In this section we characterise and argue for the existence for such an optimum. For this, we require an additional assumption on the distribution of participation preferences.

Assumption 4. *Assume $G(z)$ is such that the elasticity of take-up, $\eta(u;\sigma)$, is non-decreasing in u. This is equivalent to assuming that $G(z)/zg(z)$ is decreasing in z.*

Let $\Phi(q) \equiv \frac{s_q(q;\theta_L)}{\Delta'(q)}$ denote the ratio of the marginal change in surplus divided by the marginal change in information rent associated with varying q provided to low-risks. i.e. $\Phi(q)$ is the marginal benefit of increasing low-risks' coverage relative to its information cost.

Proposition 3. *Assume Assumptions 1, 2, 3 hold and that $G(z)$ is log-concave and satisfies Assumption 4. Then a solution to the monopoly insurance problem exists. Let this monopoly optimum be denoted by $C^m \equiv \{(u_H, q_H), (u_L, q_L)\}$. There exist threshold values, $0 \leq \bar{\sigma}_1 < \bar{\sigma}_2$, which determine the configuration of the monopoly optimum.*

1. *For $\sigma \in [0, \bar{\sigma}_1]$, C^m is such that $q_H = q^*$, $u_L = \sigma$ with $u_H = \sigma + \Delta(q_L)$ and q_L given by*

$$\Phi(q_L) - \frac{f_H}{f_L} = 0 \tag{2.1}$$

2. *For $\sigma \in (\bar{\sigma}_1, \bar{\sigma}_2]$, C^m is such that $u_H = \sigma$, $q_H = q^*$ with $u_L = \sigma - \Delta(q_L)$ and q_L given by*

$$\eta(u_L;\sigma)\frac{s(q_L;\theta_L) - u_L}{u_L} - 1 - \Phi(q_L) = 0 \tag{2.2}$$

3. *For $\sigma > \bar{\sigma}_2$, there exists a threshold proportion of high-to-low risks, $\bar{\psi} > 0$, such that if*

$\frac{f_H}{f_L} \geq \bar{\psi}$, then C^m is such that $q_H = q^*$ with u_H, u_L, q_L given by

$$u_H = u_L + \Delta(q_L) \tag{2.3}$$

$$\eta(u_L;\sigma)\frac{s(q_L;\theta_L) - u_L}{u_L} - 1 - \Phi(q_L) = 0 \tag{2.4}$$

$$\eta(u_H;\sigma)\frac{s_H^* - u_H}{u_H} - 1 + \frac{f_L}{f_H}\frac{G(u_L/\sigma)}{G(u_H/\sigma)}\Phi(q_L) = 0 \tag{2.5}$$

Point 1 of Proposition 2 says that there is a threshold level of heterogeneity below which the monopoly optimally induces full take-up by individuals of all risks. Conditional on full take-up, the monopolist maximises total surplus subject to incentive compatibility. This is done by setting q_L to trade off the marginal profit gain from low-risk individuals against the corresponding profit lost to high-risk individuals through informational rent. This trade-off is encapsulated in Condition 2.1. To extract the largest portion of total surplus, the monopolist then offers the minimum level of participant surplus that would achieve full take-up and risk-separation. This is given by $u_L = \sigma$ and $u_H = u_L + \Delta(q_L)$. Note that the monopoly optimum under full take-up in our model yields the same level of coverage provision as the optimum in the canonical setting with no heterogeneity. Having $\sigma > 0$ in this context simply shifts all individuals' surplus up by σ. For low-risk individuals, this increase maintains full take-up while for high-risk risk individuals, the shift preserves risk-separation.

Point 2 of Proposition 2 says that there is an intermediate range of heterogeneity under which the monopoly optimally induces full take-up by high-risk individuals and partial take-up by low-risks. With partial take-up by low-risks, the monopoly problem can interpreted as the following simultaneous two-part problem. First, for a given level of high-risk participant surplus, u_H, the monopoly chooses u_L and q_L to maximise total expected profit from the low-risk segment subject to incentive compatibility. Second, the monopoly chooses u_H to maximise total expected profit from the high-risk segment while accounting for the effect that u_H has on profit from the low-risk segment. We show that for $\sigma > \bar{\sigma}_1$, a monopolist optimally offers the high-risks σ, the minimum level of surplus that would achieve full take-up.

On the other hand, for any u_H, the optimal choice of u_L and q_L in the first sub-problem balances the intensive and extensive margin gains in profit from low-risks. This optimal trade-off is given by Condition 2.2. The first term captures the per-contract extensive margin gain in profit from increasing u_L holding q_L unchanged. The second term is the per-contract intensive-margin loss in profit from increasing u_L and the third term is the per-contract intensive-margin loss in profit from the corresponding decrease in q_L in order to satisfy incentive compatibility.

Point 3 of Proposition 2 says that for sufficiently high levels of heterogeneity, the monopoly optimally induces partial take-up by individuals of all risks. The condition $f_H/f_L \geq \bar{\psi}$ is a sufficient condition for an interior solution with positive take-up by both risk groups. The two-part interpretation of the monopoly problem (from Point 2) applies similarly for this case and Condition 2.4 is analogous to Condition 2.2. The difference here is that $u_H \leq \sigma$. In this regard, Condition 2.5 highlights the intensive and extensive margin gains in profit from high-risks. The first term is the per-contract extensive margin gain in profit from increasing u_H. The second term is the per-contract intensive-margin loss in profit from increasing u_H. The third term captures the intensive-margin gain in low-risk profit since increasing u_H relaxes the IC_H constraint and allows one to increase q_L. This object is divided by total demand from high-risks to give the corresponding profit gain per high-risk contract. Since $G(u_H/\sigma)$, $G(u_L/\sigma)$ are endogeneously determined, the optimality conditions in Point 3 show how with heteregeneous participation preferences and partial take-up, the monopoly accounts for the effect of its choices on the composition of risks being served.

Existence of an optimum for all σ is based on the following arguments. First the IC_H condition implies that any solution to the monopoly problem must follow one of the following three take-up regimes: full take-up by both risks, full take-up by high-risk individuals and partial take-up by low-risks or partial take-up by bothl risks. These regimes partition the solution space of the monopoly problem into three sub-spaces. Second, on each of these sub-spaces, we can find a solution to the corresponding problem to the "sub-problem" of maximising the monpolist's profit conditional on incentive compatibility and inducing take-up consistent with the sub-space. Third, for given f_H, f_L, we can find thresholds $0 \leq \bar{\sigma}_1 < \bar{\sigma}_2$ which determine where the solution to a sub-problem is also the global optimum. Note that existence does not rely on the condition, $f_H/f_L \geq \bar{\psi}$, in Point 3. Where an interior solution fails to exist, $\bar{\sigma}_2$ is the threshold above which the monopoly optimum induces zero low-risk take-up with $u_L = 0$, $q_L = 0$, $q_H = q^*$ and u_H given by $\eta(u_H;\sigma)[s^*_H - u_H]/u_H - 1 = 0$.

2.3.4 Comparative statics

Corollary 3. *Assume Assumptions 1, 2, 3 hold and $G(z)$ is log-concave and satisfies Assumption 4 The monopoly optimum denoted by $C^m \equiv \{(u_H, q_H), (u_L, q_L)\}$ is such that*

1. *For $\sigma \in [0, \bar{\sigma}_1]$, $q_L = q_L^\infty$ is invariant in σ.*
2. *For $\sigma \in (\bar{\sigma}_1, \underline{\sigma}_2]$, q_L is strictly increasing in σ.*

When the monopoly optimum has full take-up, q_L is invariant in σ. This implies the size of total surplus generated from insuring high and low-risk individuals remains constant. In this regime, changes in consumer and producer surpluses due to changes in σ derive purely from the reallocation of their respective shares of total surplus from trade. Specifically, we know from the binding incentive combatibility condition that surpluses enjoyed by high and low-risk individuals are strictly increasing in σ. Note that this part of the comparative statics results does not rely on log-concavity of G or Assumption 4. When the monopoly optimum has partial take-up by low-risks and full take-up by high-risks, coverage to low-risks, q_L, is increasing in σ. In this regime, high-risks are offered surplus equal to σ, the minimum amount required to maintain full take-up by this group. This, in turn, means that an increase in σ increases u_H and relaxes the IC_H constraint, thereby allowing q_L to be raised.

Corollary 4. *Assume Assumptions 1, 2, 3 hold and $G(z)$ is log-concave and satisfies Assumption 4 The monopoly optimum denoted by $C^m \equiv \{(u_H, q_H), (u_L, q_L)\}$ is such that*

1. *For $\sigma \in [0, \bar{\sigma}_1]$, $G(u_H/\sigma)$ and $G(u_L/\sigma)$ are invariant in σ.*
2. *For $\sigma \in (\bar{\sigma}_1, \underline{\sigma}_2]$, $G(u_H/\sigma)$ is invariant in σ while $G(u_L/\sigma)$ is strictly decreasing in σ*

By definition, $G(u_H/\sigma)$ and $G(u_L/\sigma)$ are invariant when the monopoly optimum has full take-up. When the monopoly optimum has partial take-up by low-risks and full take-up by high-risks, the low-risk take-up rate $G(u_L/\sigma)$ is decreasing in σ. As noted above, an increase in σ in this regime raises u_H and relaxes the IC_H constraint. This, in turn, allows q_L and u_L to be raised. In this regard, because raising q_L yields a positive effect on profits via intensive margin for all $q_L < q^*$, while raising u_L involves a trade-off between intensive margin profit loss and extensive margin gains, q_L always rises with σ (as noted above) and hence u_L can only increase by less than one-for-one with σ. Since the take-up rate is determined by the ratio u/σ, this implies take-up by low-risks is strictly decrasing in σ.

2.4 Conclusion

In this paper, I proposed and analyse a monopoly insurance contracting model with unobserved heterogeneity in both risk and participation preferences. This model allows us to study the effect of asymmetric information on both insurance coverage and take-up. The analysis shows that selection along either, or both, the intensive and extensive margins can arise at a monopoly optimum. The particular configuration at an optimum depends on the degree of heterogeneity in participation preferences. When there is a low degree of heterogeneity in participation preferences, there is only selection between contracts with higher risk individuals obtaining more generous coverage. When there is a high degree of heterogeneity there is simultaneously selection between contracts and in take-up. For intermediate levels of heterogeneity, an optimum features full take-up by high risk individuals and partial take-up by low-risk inviduals. As with classical Stliglitz-type models of monopoly insurance, the optimal menu of contracts offers the highest-risks full insurance coverage in all cases. On the other hand, partial coverage is provided to low-risks. In this regard, the present analysis shows that the degree of "rationing" of low-risk coverage is non-monotonic in the level of preference heterogeneity. With full take-up by all individuals, coverage levels are invariant in the degree of heterogeneity. When there is partial take-up by low-risk individuals and full take-up amongst high-risks, low-risks' coverage is increasing in heterogeneity. Finally, introducing heterogeneous participation preferences also allows us to study take-up rates of insurance. In this regard, the model implies that take-up by low-risks is strictly decreasing in the degree of participation preference heterogeneity whenever there is partial take-up by low-risk individuals and full take-up amongst high-risks.

A Proof of Lemma 3

Consider the monopoly pricing problem

$$u_i \in \max_{u \geq 0} \quad f_i G\Big(\frac{u}{\sigma}\Big)[s_i^* - u] \quad s.t. \quad u \geq 0$$

Since $G(u;\sigma) = 1$ for all $u \geq \sigma$, the solution must clearly be in $[0,\sigma]$. Among the two boundary contracts, offering $u_i = \sigma$ yields profit $\Pi(\sigma) = f_i[s_i^* - \sigma]$, while offering $u_i = 0$ yields $\Pi(0) = 0$. This means $\Pi(\sigma) \gtreqless \Pi(0) \iff \sigma \lesseqgtr s_i^*$. We now study candidate interior solutions to the problem.

A.1 Characterisation

For $u \in [0, \sigma]$, the objective has the following derivative

$$\frac{\partial \Pi}{\partial u} = f_i G'\Big(\frac{u}{\sigma}\Big)\frac{1}{\sigma}[s_i^* - u] - f_i G\Big(\frac{u}{\sigma}\Big) \tag{2.6}$$

Let $\eta(u,\sigma) \equiv \frac{u}{\sigma}\frac{G'(u/\sigma)}{G(u/\sigma)}$ denote the elasticity of take-up with respect to u. The first-order condition for an interior optimum, denoted by u_i, can be expressed as

$$f_i G\Big(\frac{u}{\sigma}\Big)\left(\eta(u_i;\sigma)\Big[\frac{s_i^*}{u_i} - 1\Big] - 1\right) = 0 \tag{2.7}$$

The surplus enjoyed by an insured individual at this allocation is

$$u_i = \frac{\eta(u_i;\sigma)}{1+\eta(u_i;\sigma)} s_i^*$$

And each contract yields profit

$$s_i^* - u_i = \frac{1}{1+\eta(u_i;\sigma)} s_i^*$$

A.2 Sufficiency conditions

The second derivative with respect to u is

$$\begin{aligned}
\frac{\partial^2 \Pi}{\partial u^2} =& f_i G''\Big(\frac{u}{\sigma}\Big)\frac{1}{\sigma^2}[s_i^* - u] - 2 f_i G'\Big(\frac{u}{\sigma}\Big)\frac{1}{\sigma} \\
=& f_i G\Big(\frac{u}{\sigma}\Big)\frac{G'(u/\sigma)}{G(u/\sigma)}\frac{1}{\sigma}\left(\frac{G''(u/\sigma)}{G'(u/\sigma)}\frac{1}{\sigma}[s_i^* - u] - 2\right) \\
\le& f_i G\Big(\frac{u}{\sigma}\Big)\frac{G'(u/\sigma)}{G(u/\sigma)}\frac{1}{\sigma}\left(\frac{G'(u/\sigma)}{G(u/\sigma)}\frac{1}{\sigma}[s_i^* - u] - 2\right) \\
\le& f_i G\Big(\frac{u}{\sigma}\Big)\frac{G'(u/\sigma)}{G(u/\sigma)}\frac{1}{\sigma}\left(\frac{G'(1)}{G(1)}\frac{1}{\sigma}[s_i^* - \sigma] - 1 + \frac{G'(1)}{G(1)} - 1\right) \\
\le& f_i G\Big(\frac{u}{\sigma}\Big)\frac{G'(u/\sigma)}{G(u/\sigma)}\frac{1}{\sigma}\left(\frac{G'(1)}{G(1)}\frac{1}{\sigma}[s_i^* - \sigma] - 1\right) \le 0
\end{aligned}$$

Since total expected profit is negative for all $u \ge s_i^*$ and any such contract is dominated by offering $u = 0$, we can restrict attention to $u \in [0, \min\{\sigma,\ s_i^*\}]$. The first inequality obtains because log-concavity of G implies $G''(u/\sigma)/G'(u/\sigma) \le G'(u/\sigma)/G(u/\sigma)$. The second inequality arises since we consider $\sigma < s^*$ and because log-concavity of G implies $G'(u/\sigma)/G(u/\sigma)$ is decreasing in u and hence is maximum at $G'(1)/G(1)$. The third inequality arises by definition of the derivative of a distribution function. The last inequality obtains from the Kuhn-Tucker first-order condition with the constraint $u \leq \sigma$.

A.3 Comparative statics

The second partial derivative with respect to σ is

$$\begin{aligned}\frac{\partial^2 \Pi}{\partial \sigma \partial u} &= -\frac{u}{\sigma} G''\Big(\frac{u}{\sigma}\Big)\frac{1}{\sigma^2}[s_i^* - u] - \frac{1}{\sigma} G'\Big(\frac{u}{\sigma}\Big)\frac{1}{\sigma}[s_i^* - u] + \frac{u}{\sigma} G'\Big(\frac{u}{\sigma}\Big)\frac{1}{\sigma} \\ &= -G\Big(\frac{u}{\sigma}\Big)\frac{1}{\sigma}\left(1 + \frac{u}{\sigma}\left[\frac{G''(u/\sigma)}{G'(u/\sigma)} - \frac{G'(u/\sigma)}{G(u/\sigma)}\right]\right) \end{aligned} \tag{2.8}$$

By the implicit function theorem,

$$\frac{du_i}{d\sigma} = \frac{1 + \frac{u_i}{\sigma}\left[\frac{G''(u_i/\sigma)}{G'(u_i/\sigma)} - \frac{G'(u_i/\sigma)}{G(u_i/\sigma)}\right]}{\frac{G''(u_i/\sigma)}{G'(u_i/\sigma)} - 2\frac{G'(u_i/\sigma)}{G(u_i/\sigma)}} \tag{2.9}$$

First, note that

$$\frac{\sigma}{u_i}\frac{du_i}{d\sigma} = 1 + \frac{\frac{\sigma}{u_i} + \frac{G'(u_i/\sigma)}{G(u_i/\sigma)}}{\frac{G''(u_i/\sigma)}{G'(u_i/\sigma)} - 2\frac{G'(u_i/\sigma)}{G(u_i/\sigma)}} < 1 \tag{2.10}$$

This implies $G(u_i/\sigma)$ is strictly decreasing in σ. Next, note that differentiating the the definition of $\eta(u;\sigma)$ yields

$$\frac{d\eta(u;\sigma)}{du} = \frac{G'(u/\sigma)}{G(u/\sigma)}\frac{1}{\sigma}\left(1 + \frac{u}{\sigma}\left[\frac{G''(u/\sigma)}{G'(u/\sigma)} - \frac{G'(u/\sigma)}{G(u/\sigma)}\right]\right) \tag{2.11}$$

Therefore, $\frac{du_i}{d\sigma} \gtrless 0 \iff \frac{d\eta(u/\sigma)}{du} \lessgtr 0$.

Optimal partial take-up

The insurer's profit under candidate interior solution is

$$\Pi(u_i) = f_i G(u_i;\sigma)\frac{1}{1+\eta(u_i;\sigma)} s_i^*$$

Clearly, $\Pi(u_i) > 0 = \Pi(0)$ so we are left to compare the interior solution against the boundary case with $u_i = \sigma$. The difference in total expected profit is

$$\Pi(\sigma) - \Pi(u_i) = f_i[s_i^* - \sigma] - f_i G(u_i;\sigma)[s_i^* - u_i] = f_i[s_i^* - \sigma] - f_i G(u_i;\sigma)\frac{1}{1+\eta(u_i;\sigma)}s_i^*$$

Define

$$\Xi(\sigma) \equiv s_i^* - \sigma - G(u_i;\sigma)\frac{1}{1+\eta(u_i;\sigma)}s_i^* = \frac{\eta(u_i;\sigma) + 1 - G(u_i;\sigma)}{1+\eta(u_i;\sigma)}s_i^* - \sigma \tag{2.12}$$

So $\Pi(\sigma) - \Pi(u_i) \gtrless 0 \iff \Xi(\sigma) \gtrless 0$. We want to show that there exists a unique threshold, $\bar{\sigma}$, such that $\Xi(\bar{\sigma}) = 0$ and $\Xi(\sigma) < 0$ for all $\sigma > \bar{\sigma}$.

First, evaluating the optimality condition for an interior solution at $u_i = \sigma$, we have

$$u_i = \sigma \iff \sigma = \frac{\eta(\sigma;\sigma)}{1+\eta(\sigma;\sigma)}s_i^*$$

Note that $\eta(\sigma;\sigma) = g(1)$, where $g(z)$ is the density of the (normalised) random variable $z = u/\sigma$. For a given distribution $g(1)$, and hence $\eta(\sigma;\sigma)$, is invariant in σ. Thus, there is a unique positive and constant value of σ at which the candidate interior solution is $u_i = \sigma$. Denote this value by $\bar{\sigma}_i$ and note that $\Xi(\bar{\sigma}_i) = 0$.

Next, differentiating $\Xi(\sigma)$ by the envelope condition gives

$$\begin{aligned}
\Xi'(\sigma) &= -1 - \frac{1}{f_i}\frac{\partial \Pi(u_i)}{\partial \sigma} \\
&= -1 - \frac{\partial G(u_i;\sigma)}{\partial \sigma}[s_i^* - u_i] \\
&= -1 + \eta(u_i;\sigma)\frac{1}{\sigma}G(u_i;\sigma)[s_i^* - u_i] \\
&= -1 + G(u_i;\sigma)\frac{\eta(u_i;\sigma)}{1+\eta(u_i;\sigma)}s_i^*\frac{1}{\sigma}
\end{aligned}$$

Using $u_i = \frac{\eta(u_i;\sigma)}{1+\eta(u_i;\sigma)}s_i^*$, we have

$$G(u_i;\sigma)\frac{\eta(u_i;\sigma)}{1+\eta(u_i;\sigma)}s_i^*\frac{1}{\sigma} = G(u_i;\sigma)\frac{u_i}{\sigma} \leq G(u_i;\sigma) \leq 1$$

So $\Xi'(\sigma) \leq 0$ with $\Xi'(\sigma) = 0$ at $\sigma = \bar{\sigma}_i$. Moreover, since we have argued that $G(u_i;\sigma)$ is strictly decreasing in σ, we have $\Xi'(\sigma) < 0$ for all $\sigma > \bar{\sigma}_i$.

In sum, we have shown that $\Pi(\sigma) = \Pi(u_i)$ at $\sigma = \bar{\sigma}_i$ and $\Pi(\sigma) < \Pi(u_i)$ when $\sigma > \bar{\sigma}_i$.

Assume that the global concavity condition is satisfied, then our preceding analysis shows that there exists a unique solution to the monopoly problem given by $u_i = \sigma$ for all $\sigma \in [0, \bar{\sigma}_i]$ and $u_i = \frac{\eta(u_i;\sigma)}{1+\eta(u_i;\sigma)} s_i^*$ for all $\sigma > \bar{\sigma}_i$.

B General properties

We first generalise the well-known result monotonicity property of incentive compatible allocations to our setting. This is a generalisation because we show that monotonicity is implied along both the intensive and extensive margins

Lemma 4. *Any incentive compatible allocation must have $q_H \geq q_L$ and $G(u_H/\sigma) \geq G(u_L/\sigma)$.*

Proof. Combining (IC_H) and (IC_L) yields $\Delta(q_H) \geq u_H - u_L \geq \Delta(q_L)$ so $q_H \geq q_L$. Next, since $\Delta(q_L) \geq 0$ for all $q_L \geq 0$, any allocation which satisfies (IC_H) has $u_H \geq u_L$. Furthermore, when θ and y are independently distributed, $u_H \geq u_L$ implies $1 \geq G(u_H) \geq G(u_L)$. Specifically, when $u_H \geq u_L \geq \sigma$, we have $G(u_H) = G(u_H) = 1$ and when $u_H \geq \sigma > u_L$, we have $G(u_H) = 1 > G(u_H)$. $\square$

Note that Lemma 4 implies that the constraints $q_H \geq 0$ and $u_H \geq 0$ are redundant. Furthermore, because $u_H \geq u_L$, the monopoly optimum must be in one of the following three subspaces.

$$\begin{aligned}
\mathcal{C}_1 &\equiv \Big\{(u_H, q_H, u_L, q_L) \in \mathbb{R}_+^4 : u_H \geq \sigma,\ u_L \geq \sigma,\ q_L \geq 0,\ (IC_H),\ (IC_L)\Big\} \\
\mathcal{C}_2 &\equiv \Big\{(u_H, q_H, u_L, q_L) \in \mathbb{R}_+^4 : u_H \geq \sigma,\ \sigma \geq u_L \geq 0,\ q_L \geq 0,\ (IC_H),\ (IC_L)\Big\} \\
\mathcal{C}_3 &\equiv \Big\{(u_H, q_H, u_L, q_L) \in \mathbb{R}_+^4 : \sigma \geq u_H \geq 0,\ \sigma \geq u_L \geq 0,\ q_L \geq 0,\ (IC_H),\ (IC_L)\Big\}
\end{aligned}$$

Let $\overline{\mathcal{C}}_1$ denote the subspace of $\mathcal{C}_1$ where $u_L = \sigma$ is binding, $\overline{\mathcal{C}}_2$ the subspace of $\mathcal{C}_2$ where $u_H = \sigma$ is binding and $\overline{\mathcal{C}}_3$ the subspace of $\mathcal{C}_3$ where $u_L = 0$ is binding. We hereafter call these boundary allocations on each of their respective subspaces.

We next establish the following well-known result that both incentive compatibility constraints cannot be simultaneously binding.

Lemma 5. *A solution to the monopoly problem cannot simultaneously have $\lambda_H > 0$ and $\lambda_L > 0$.*

Proof. Suppose $\lambda_H > 0$ and $\lambda_L > 0$. This implies IC_H and IC_L hold with equality which, in turn, implies $\Delta(q_L) = u_H - u_L = \Delta(q_H)$ and $q_H = q_L$. At the same time, the derivative of the objective with respect to q_H reads $f_H G(u_{H;\sigma})s_q(q_H;\theta_H) = -\lambda_L \Delta'(q_H) < 0$ so $q_H > q^*$. First take $\varphi > 0$. Then we have an immediate contradiction since the binding constraint implies $q_L = 0 < q^* < q_H$. Next take $\varphi = 0$. Then the derivative of the objective with respect to q_L reads $f_L G(u_L/\sigma)s_q(q_L;\theta_L) = \lambda_H \Delta'(q_L) > 0$ and hence $q_L < q^*$. Thus, $\lambda_H > 0$ and $\lambda_L > 0$ implies $q_H > q^* > q_L$, a contradiction. □

Next, note that for each $\mathcal{C}_r$, $r = 1,2,3$, we can define the sub-problem of maximising the monopoly's total expected profit over the particular sub-space. Each of these problems corresponds to Problem $\mathcal{P}^m$ with additional "participation" constraints. Specifically, sub-problem $\mathcal{P}_1^m$ on $\mathcal{C}_1$, includes the constraint $u_L \geq \sigma$, sub-problem $\mathcal{P}_2^m$ on $\mathcal{C}_2$ adds the constraints $u_H \geq \sigma$ and $u_L \leq \sigma$. Finally, sub-problem $\mathcal{P}_3^m$ on $\mathcal{C}_3$, we includes the constraint $u_H \leq \sigma$. To fix terminology, in our subsequent exposition, we call solutions in which the participation constraints are non-binding "interior solutions" and refer to solutions with binding participation constraints "boundary solutions".

The next lemma establishes the familiar result that *interior solutions* of the insurer's problem must have the IC_H constraint binding.

Lemma 6. *Assume that $G'(z)/G(z)$ is non-increasing. If a solution to the monopoly problem is in $\overline{\mathcal{C}}_1$, $\underline{\mathcal{C}}_2$ or $\underline{\mathcal{C}}_3$, then this solution must be such that $\lambda_H > 0$.*

Proof. Any interior solution to the sub-problems must satisfy the necessary F.O.C.s

$$f_H G'\Big(\frac{u_H}{\sigma}\Big)\frac{1}{\sigma}[s(q_H;\theta_H) - u_H] - f_H G\Big(\frac{u_H}{\sigma}\Big) + \lambda_H - \lambda_L = 0 \tag{2.13}$$

$$f_L G'\Big(\frac{u_L}{\sigma}\Big)\frac{1}{\sigma}[s(q_L;\theta_L) - u_L] - f_L G\Big(\frac{u_L}{\sigma}\Big) - \lambda_H + \lambda_L = 0 \tag{2.14}$$

$$f_H G\Big(\frac{u_H}{\sigma}\Big)s_q(q_H;\theta_H) + \lambda_L \Delta'(q_H) = 0 \tag{2.15}$$

$$f_L G\Big(\frac{u_L}{\sigma}\Big)s_q(q_L;\theta_L) - \lambda_H \Delta'(q_L) + \varphi = 0 \tag{2.16}$$

$$\lambda_H[u_H - u_L - \Delta(q_L)] = 0,\ \lambda_H \geq 0,\ u_H - u_L - \Delta(q_L) \geq 0$$

$$\lambda_L[u_L - u_H + \Delta(q_H)] = 0,\ \lambda_L \geq 0,\ u_L - u_H + \Delta(q_H) \geq 0$$

$$\varphi q_L = 0,\ \varphi \geq 0,\ q_L \geq 0$$

Suppose, for a contradiction, that $\lambda_H = 0$ at the solution to this system. For the case where $\hat{C}$ is in $\mathcal{C}_1$ or $\mathcal{C}_2$, we have $\frac{\partial G(u_H;\sigma)}{\partial u} = 0$ so (2.13) reads $-f_H - \lambda_L < 0$, a contradiction. We are thus left to consider $\hat{C} \in \mathcal{C}_3$. In this case, (2.13) and (2.14) imply

$$[s(q_H;\theta_H) - u_H] - [s(q_L;\theta_L) - u_L] \geq \frac{G(u_H\sigma)}{G'(u_H/\sigma)}\sigma - \frac{G(u_L/\sigma)}{G'(u_L/\sigma)}\sigma > 0$$

Next, note that when $u_L > 0$, (2.16) implies we have $q_L = q^*$ while (2.15) implies $q_H \geq q^*$. Adding $s(q_H;\theta_H)$ and subtracting $s(q_L;\theta_L)$ to both sides of (IC_L), we have

$$\begin{aligned}[s(q_H;\theta_H) - u_H] - [s(q^*;\theta_L) - u_L] =& s(q_H;\theta_H) - s(q^*;\theta_L) - \Delta(q^*) \\ =& -c(q_H;\theta_H) + c(q^*;\theta_L) \leq -c(q^*;\theta_H) + c(q^*;\theta_L) < 0\end{aligned}$$

We thus obtain our desired contradiction. □

Lemmas 5 and 6 imply that any interior solution in $\mathcal{C}_1$, $\mathcal{C}_2$ or $\mathcal{C}_3$ must be such that $\lambda_H > 0$ and $\lambda_L = 0$. This, in turn, implies $q_H = q^*$ and $u_H = u_L + \Delta(q_L)$.

A classic result is that a monopolist's optimal pricing schedule will leave no rent to the "lowest type" beyond what is necessary to induce participation. The next lemma generalises this result to the case with heterogeneous participation preferences.

Lemma 7. *A solution to the monopoly problem is such that $u_L \leq \sigma$.*

Proof. Suppose, for a contradiction, that $\hat{C} \equiv (\hat{u}_H, \hat{q}_H, \hat{u}_L, \hat{q}_L)$ where $\hat{C} \in \mathcal{C}_1$ and $\hat{u}_L > \sigma$ is a solution to Problem $\mathcal{P}^m$. Then by (IC_H), we have $\hat{u}_H > \sigma$. Hence we have $G(\hat{u}_H;\sigma) = G(\hat{u}_L;\sigma) = 1$ with their respective derivatives with respect to u equal to zero. Moreover, since $\hat{C}$ solves Problem $\mathcal{P}^m$ and is an interior point of the subspace $\mathcal{C}_1$, it must satisfy the following necessary first-order conditions

$$-f_H + \lambda_H - \lambda_L = 0 \tag{2.17}$$

$$-f_L - \lambda_H + \lambda_L = 0 \tag{2.18}$$

$$f_H s_q(q_H;\theta_H) + \lambda_L \Delta'(q_H) = 0$$

$$f_L s_q(q_L;\theta_L) - \lambda_H \Delta'(q_L) + \varphi = 0$$

$$\lambda_H[u_H - u_L - \Delta(q_L)] = 0,\ \lambda_H \geq 0,\ u_H - u_L - \Delta(q_L) \geq 0$$

$$\lambda_L[u_L - u_H + \Delta(q_H)] = 0,\ \lambda_L \geq 0,\ u_L - u_H + \Delta(q_H) \geq 0$$

$$\varphi q_L = 0,\ \varphi \geq 0,\ q_L \geq 0$$

Suppose $\lambda_H > 0$, then by Lemma 5, $\lambda_L = 0$. However, this violates (2.18). Next, suppose $\lambda_L > 0$, then by Lemma 5, $\lambda_H = 0$ but this violates (2.17). Finally, $\lambda_H = \lambda_L = 0$ violates both (2.17) and (2.18). Therefore, $\hat{C}$ cannot be a solution to Problem $\mathcal{P}^m$. □

Thus, with heterogeneous participation preferences, the "lowest type" never obtains more surplus than is necessary to induce full-take up and indeed can obtain less. Moreover, Lemma 7 implies we can omit interior solutions on $\mathcal{C}_1$ from the remainder of our analysis.

C Solutions on each subspace

C.1 Boundary solution on C_1

As argued above, any solution to the monopoly problem on C_1 must be a boundary allocation with $u_L = \sigma$. Moreover, by the same arguments as above, any such solution must have $\lambda_H > 0$, $\lambda_L = 0$, $q_H = q^*$ and $u_H = u_L + \Delta(q_L)$. Given the above points, a solution to the monopoly problem on C_1 equivalently solves

$$\max_q \quad f_H\Big[s_H^* - \sigma - \Delta(q)\Big] + f_L\Big[s(q;\theta_L) - \sigma\Big] \quad s.t. \quad q \geq 0$$

Characterisation

Let $\Phi(q) \equiv \frac{s_q(q;\theta_L)}{\Delta'(q)}$ and note that $\Phi'(q) < 0$ for all $q \geq 0$, $\max\{\Phi(q)\} = \Phi(0)$ and $\Phi(q^*) = 0$. A solution to this problem must satisfy the following necessary F.O.C.

$$\begin{aligned} -f_H + f_L\Phi(q_L) + \varphi &= 0 \\ \varphi q_L = 0,\ \varphi \geq 0,\ q_L &\geq 0 \end{aligned} \tag{2.19}$$

Suppose $\varphi = 0$ and $q_L > 0$. Because $\Phi(q)$ is strictly decreasing in q, (2.19) implies that q_L is strictly decreasing in $\frac{f_H}{f_L}$. This implies that the constraint $q_L \geq 0$ is binding if and only if $\frac{f_H}{f_L} > \Phi(0)$.

Sufficiency

As noted above, $\Phi'(q) < 0$ for all $q \geq 0$, therefore the objective is strictly concave for all $q \geq 0$ and hence (2.19) characterises the unique solution to sub-problem $\mathcal{P}_1^m$.

Summary

- For $\frac{f_H}{f_L} \le \Phi(0)$, the solution to the problem is $C_1 \equiv \Big\{(u_{H,1}, q_{H,1}), (u_{L,1}, q_{L,1})\Big\} = \Big\{\Big(\sigma + \Delta(q_L^\infty), q^*\Big), \Big(\sigma, q_L^\infty\Big)\Big\}$ and the total expected profit at this allocation is

$$\Pi(C_1) = f_H[s_H^* - \sigma - \Delta(q_L^\infty)] + f_L\Big[s(q_L^\infty; \theta_L) - \sigma\Big] \tag{2.20}$$

- For $\frac{f_H}{f_L} > \Phi(0)$, the solution is $C_1 \equiv \Big\{(u_{H,1}, q_{H,1}), (u_{L,1}, q_{L,1})\Big\} = \Big\{\Big(\sigma, q^*\Big), \Big(\sigma, 0\Big)\Big\}$ and the total expected profit at this allocation is

$$\Pi(C_1) = f_H[s_H^* - \sigma] - f_L\sigma \tag{2.21}$$

Remark that when $\sigma = 0$, the solution described above becomes

- For $\frac{f_H}{f_L} \le \Phi(0)$, the solution to the problem is $C_1 = \{(\Delta(q_L^\infty), q^*), (0, q_L^\infty)\}$ and the total expected profit at this allocation is $\Pi(C_1) = f_H[s_H^* - \Delta(q_L^\infty)] + f_L s(q_L^\infty; \theta_L)$
- For $\frac{f_H}{f_L} > \Phi(0)$, the solution is $C_1 = \{(0, q^*), (0, 0)\}$ and the total expected profit at this allocation is $\Pi(C_1) = f_H s_H^*$

This means the solution to the full take-up sub-problem reduces to the monopoly optimum in canonical monopoly insurance problem with no participation preferences.

C.2 Interior solution on $\mathcal{C}_2$

Because we know that IC_H is binding and IC_L is non-binding at any interior solution to the monopoly problem, we can reformulate the monopoly problem on subspace $\mathcal{C}_2$ as the following simultaneous sub-problems

$$\max_{U} \quad f_H[s_H^* - U] + f_L \pi(U) \tag{$\mathcal{P}_2^{outer}$}$$

$$\textit{where} \quad \pi(U) \equiv \max_{(u,q)\in[0,\sigma]\times\mathbb{R}_+} \quad G(u;\sigma)[s(q;\theta_L) - u] \quad \textit{s.t.} \quad u + \Delta(q) \le U \tag{$\mathcal{P}^{inner}$}$$

We can think of the interior problem as maximising total expected profit from the low-risk segment by choosing (u, q) subject to an "incentive budget", U. In turn, the outer problem maximises total expected profit from the high-risk segment by choosing U while accounting for the fact that U affects the incentive budget of the inner problem.

Inner Problem

A solution to Problem $\mathcal{P}_2^{inner}$ must satisfy the F.O.C.

$$0 = \frac{\partial \pi}{\partial u} = G'\Big(\frac{u}{\sigma}\Big)\frac{1}{\sigma}\Big[s\Big(q_L(U,u);\theta_L\Big) - u\Big] - G\Big(\frac{u}{\sigma}\Big) - G\Big(\frac{u}{\sigma}\Big)\Phi\Big(q_L(U,u)\Big) \tag{2.22}$$

Sufficiency The second derivative is

$$\frac{\partial^2 \pi}{\partial u^2} = G''\Big(\frac{u}{\sigma}\Big)\frac{1}{\sigma^2}\Big[s\Big(q_L(U,u);\theta_L\Big) - u\Big] - 2G'\Big(\frac{u}{\sigma}\Big)\frac{1}{\sigma} - 2G'\Big(\frac{u}{\sigma}\Big)\frac{1}{\sigma}\Phi\Big(q_L(U,u)\Big) + G\Big(\frac{u}{\sigma}\Big)\frac{\Phi'\Big(q_L(U,u)\Big)}{\Delta'\Big(q_L(U,u)\Big)}$$

By log-concavity of $G(z)$,

$$\begin{aligned}
&G''\Big(\frac{u}{\sigma}\Big)\frac{1}{\sigma^2}\Big[s\Big(q_L(U,u);\theta_L\Big) - u\Big] - 2G'\Big(\frac{u}{\sigma}\Big)\frac{1}{\sigma} - 2G'\Big(\frac{u}{\sigma}\Big)\frac{1}{\sigma}\Phi\Big(q_L(U,u)\Big) + G\Big(\frac{u}{\sigma}\Big)\frac{\Phi'\Big(q_L(U,u)\Big)}{\Delta'\Big(q_L(U,u)\Big)} \\
&\leq G'\Big(\frac{u}{\sigma}\Big)\frac{1}{\sigma}\left(\frac{G''(u/\sigma)}{G'(u/\sigma)}\frac{1}{\sigma}\Big[s\Big(q_L(U,u);\theta_L\Big) - u\Big] - 2 - 2\Phi\Big(q_L(U,u)\Big)\right) \\
&\leq G'\Big(\frac{u}{\sigma}\Big)\frac{1}{\sigma}\left(\frac{G'(u/\sigma)}{G(u/\sigma)}\frac{1}{\sigma}\Big[s\Big(q_L(U,u);\theta_L\Big) - u\Big] - 2\right) < 0
\end{aligned}$$

so the global concavity condition is satisfied.

Comparative statics The second partial derivative of the objective with respect to U is

$$\frac{\partial^2 \pi}{\partial U \partial u} = G'\Big(\frac{u}{\sigma}\Big)\frac{1}{\sigma}\Phi\Big(q_L(U,u)\Big) - G\Big(\frac{u}{\sigma}\Big)\frac{\Phi'\Big(q_L(U,u)\Big)}{\Delta'\Big(q_L(U,u)\Big)} > 0$$

Hence u is strictly increasing in U. The second partial derivative with respect to σ is

$$\begin{aligned}
\frac{\partial^2 \pi}{\partial \sigma \partial u} &= -\frac{u_L}{\sigma^2}G''\Big(\frac{u}{\sigma}\Big)\frac{1}{\sigma}\Big[s\Big(q_L(U,u);\theta_L\Big) - u\Big] - \frac{1}{\sigma}G'\Big(\frac{u}{\sigma}\Big)\frac{1}{\sigma}\Big[s\Big(q_L(U,u);\theta_L\Big) - u\Big] + \frac{u_L}{\sigma^2}G'\Big(\frac{u}{\sigma}\Big)\Big[1 + \Phi\Big(q_L(U,u)\Big)\Big] \\
&= -f_L G\Big(\frac{u}{\sigma}\Big)\frac{1}{\sigma}\Big[1 + \Phi\Big(q_L(U,u)\Big)\Big]\left(1 + \frac{u}{\sigma}\Big[\frac{G''(u/\sigma)}{G'(u/\sigma)} - \frac{G'(u/\sigma)}{G(u/\sigma)}\Big]\right)
\end{aligned}$$

Since

$$\frac{d\eta(u;\sigma)}{du} = \frac{G'(u/\sigma)}{G(u/\sigma)}\frac{1}{\sigma}\left(1+\frac{u}{\sigma}\left[\frac{G''(u/\sigma)}{G'(u/\sigma)}-\frac{G'(u/\sigma)}{G(u/\sigma)}\right]\right)$$

We have $\frac{du}{d\sigma} \gtrless 0 \iff \frac{d}{du}\eta(u/\sigma) \lessgtr 0$.

Outer Problem

A solution to Problem $\mathcal{P}_2^{outer}$ must satisfy the F.O.C.

$$-f_H + f_L\pi'(U) = 0 \quad where \quad \pi'(U) = G(u_L/\sigma)\Phi(q_L(U,u_L)) \tag{2.23}$$

First, note that $\pi'(U)$ is bounded above by $\Phi(0)$. Therefore, if $f_H/f_L > \Phi(0)$, $-f_H+f_L\pi'(U) < 0$ for all U and hence there is no interior solution to Problem $\mathcal{P}_2^{outer}$.

For the remainder of the section, suppose $f_H/f_L \leq \Phi(0)$. Note that at $\hat{\sigma}_L$ given by $G'(1)[s(q_L^\infty;\theta_L) - \hat{\sigma}_L + \Delta(q_L^\infty)] - 1 - \Phi(q_L^\infty) = 0$, $U = \hat{\sigma}_L + \Delta(q_L^\infty)$ solves F.O.C. (2.23). Therefore, there is an interior solution to the monopoly problem on $\mathcal{C}_2$ at the specific point where $\sigma = \hat{\sigma}_L$.

Next, differentiating $\pi'(U)$ we have

$$\begin{aligned}\frac{d\pi'(U)}{d\sigma} &= \frac{\partial\pi'(U)}{\partial\sigma} + \frac{\partial\pi'(U)}{\partial u_L}\frac{du_L}{d\sigma}\\ &= -\frac{u_L}{\sigma}G'\Big(\frac{u_L}{\sigma}\Big)\frac{1}{\sigma}\Phi(q_L) + \left(G'\Big(\frac{u_L}{\sigma}\Big)\frac{1}{\sigma}\Phi(q_L) - G\Big(\frac{u_L}{\sigma}\Big)\frac{\Phi'(q_L)}{\Delta'(q_L)}\right)\frac{du_L}{d\sigma}\end{aligned}$$

Under Assumption 4, $\eta(u;\sigma)$ is increasing in u so $\frac{d\pi'(U)}{d\sigma} < 0$. Because $\pi'(U)$ is strictly decreasing in σ, F.O.C. (2.23) is satisfied if and only if $\sigma = \hat{\sigma}_L$. Indeed, we have $-f_H + \pi'(U) < 0$ for all $\sigma > \hat{\sigma}_L$ and $-f_H + \pi'(U) > 0$ for all $\sigma < \hat{\sigma}_L$.

In sum, the only the interior solution on $\mathcal{C}_2$ occurs at the "knife-edge case" where $\sigma = \hat{\sigma}_L$. At this specific point, the solution to the monopoly problem is given by $u_L = \hat{\sigma}_L$

$$-f_H + f_L\Phi(q_L^\infty) = 0 \tag{2.24}$$

$$G'(1)[s(q_L^\infty;\theta_L) - \hat{\sigma}_L] - 1 - \Phi(q_L^\infty) = 0 \tag{2.25}$$

$$u_H - \hat{\sigma}_L - \Delta(q_L^\infty) = 0 \tag{2.26}$$

C.3 Boundary solution on $\mathcal{C}_2$

Now consider the sub-problem of maximising the monopolist's total profit subject to $u_H = \sigma$. This problem reads

$$\begin{aligned}
\max_{\{(u_i,q_i)\}_{i=H,L}} \quad & f_H\Big[s(q_H;\theta_H) - \sigma\Big] + f_L G\Big(\frac{u_L}{\sigma}\Big)\Big[s(q_L;\theta_L) - u_L\Big] \\
s.t. \quad & u_H \geq u_L + \Delta(q_L) \\
& u_L \geq u_H - \Delta(q_H) \\
& q_L \geq 0 \\
& u_L \geq 0
\end{aligned}$$

Characterisation

A solution to the problem must satisfy the following necessary F.O.C.s

$$f_L G'\Big(\frac{u_L}{\sigma}\Big)\frac{1}{\sigma}\Big[s(q_L;\theta_L) - u_L\Big] - f_L G\Big(\frac{u_L}{\sigma}\Big) - \lambda_H + \lambda_L + \bar{\mu} = 0 \tag{2.27}$$

$$f_H s_q(q_H;\theta_H) + \lambda_L \Delta'(q_H) = 0 \tag{2.28}$$

$$f_L G\Big(\frac{u_L}{\sigma}\Big) s_q(q_L;\theta_L) - \lambda_H \Delta'(q_L) + \varphi = 0 \tag{2.29}$$

$$\lambda_H[\sigma - u_L - \Delta(q_L)] = 0,\ \lambda_H \geq 0,\ \sigma - u_L - \Delta(q_L) \geq 0$$

$$\lambda_L[u_L - \sigma + \Delta(q_H)] = 0,\ \lambda_L \geq 0,\ u_L - \sigma + \Delta(q_H) \geq 0$$

$$\varphi q_L = 0,\ \varphi \geq 0,\ q_L \geq 0$$

$$\bar{\mu} u_L = 0,\ \bar{\mu} \geq 0,\ u_L \geq 0$$

First, note that it is possible to for a solution to this sub-problem to have $\lambda_H = 0$. Nevertheless, we can reduce the set of possible binding constraints by arguing that IC_L cannot be binding at a solution to this sub-problem.

Lemma 8. *Assume $G(z)$ log-concave. A boundary solution to the monopoly problem on $\mathcal{C}_2$ must be such that $\lambda_L = 0$*

Proof. Suppose, for a contradiction, that a boundary solution on $\mathcal{C}_2$ has $\lambda_L > 0$. By Lemma 5 we must have $\lambda_H = 0$. Because $\lambda_H = 0$, (2.29) reads $f_L G(u_L;\sigma) s_q(q_L;\theta_L) + \varphi = 0$. This condition is satisfied only if $\varphi = 0$ and either $u_L = 0$ or $q_L = q^*$. Consider the first case where $u_L = 0$, then (2.27) reads $f_L \frac{\partial G(0;\sigma)}{\partial u} s(q_L;\theta_L) + \lambda_L + \bar{\mu} > 0$, a contradiction. We are thus left with the case where $q_L = q^*$. Substituting into (2.27) yields

$$f_L G\Big(\frac{u_L}{\sigma}\Big)\left(\frac{G'(u_L/\sigma)}{G(u_L/\sigma)}\frac{1}{\sigma}[s_L^* - u_L] - 1\right) = -\lambda_L < 0 \iff s_L^* - u_L < \frac{G(u_L/\sigma)}{G'(u_L/\sigma)}\sigma$$

Next, note that since our hypothetical solution is in $\mathcal{C}_2$, it must be that

$$f_H G'(1)\frac{1}{\sigma}[s(q_H;\theta_H) - \sigma] - f_H \geq 0 \iff s(q_H;\theta_H) - \sigma \geq \frac{1}{G'(1)}\sigma$$

This condition says that the monopolist's total expected profit cannot be increased when u_H is lowered from $u_H = \sigma$. Together, these two inequalities yield

$$[s(q_H;\theta_H) - \sigma] - [s_L^* - u_L] > \sigma\left[\frac{G(1)}{G'(1)} - \frac{G(u_L/\sigma)}{G'(u_L/\sigma)}\right]$$

Note that $G(z)$ log-concave implies $G'(z)/G(z)$ is decreasing in z. This means $[s_H^* - u_H] - [s_L^* - u_L] > 0$. On the other hand, adding $s(q_H;\theta_H)$ to and subtracting s_L^* from both sides of (IC_L) yields

$$[s(q_H;\theta_H) - \sigma] - [s_L^* - u_L] = s(q_H;\theta_H) - s_L^* - \Delta(q_H) = s(q_H;\theta_L) - s(q^*;\theta_L) - \Big[c(q_H;\theta_H) - c(q_L;\theta_L)\Big] < 0$$

We thus have our desired contradiction. □

Given the previous result, the necessary F.O.C.s reduce to

$$q_H = q^*$$

$$f_L G'\Big(\frac{u_L}{\sigma}\Big)\frac{1}{\sigma}\Big[s(q_L;\theta_L) - u_L\Big] - f_L G\Big(\frac{u_L}{\sigma}\Big) - \lambda_H + \bar{\mu} = 0 \tag{2.30}$$

$$f_L G\Big(\frac{u_L}{\sigma}\Big)\Phi(q_L) - \lambda_H + \varphi = 0 \tag{2.31}$$

$$\lambda_H[\sigma - u_L - \Delta(q_L)] = 0,\ \lambda_H \geq 0,\ \sigma - u_L - \Delta(q_L) \geq 0$$

$$\varphi q_L = 0,\ \varphi \geq 0,\ q_L \geq 0$$

$$\bar{\mu} u_L = 0,\ \bar{\mu} \geq 0,\ u_L \geq 0$$

Lemma 9. *The unique boundary solution to the monopoly problem on $\mathcal{C}_2$ when $\sigma = 0$ is $\bar{C}_2 = \{(\sigma, q^*), (0,0)\}$. Moreover, at this solution, the IC_H constraint, the non-negative provision constraint, $q_L \geq 0$, and the non-negative participant surplus constraint, $u_L \geq 0$ are binding.*

Proof. Suppose $q_L = 0$. Then (**??**) reads $-u_L f_L G'(u_L/\sigma)\frac{1}{\sigma} - f_L G(u_L;\sigma) - \lambda_H + \bar{\mu} = 0$. This equation holds if and only if $u_L = 0$ and either $\lambda_H = 0$ and $\bar{\mu} = 0$ *or* $\lambda_H > 0$ and $\bar{\mu} > 0$. Suppose $\lambda_H = 0$ and $\bar{\mu} = 0$ then we must have $\varphi = 0$. However, because IC_H is not binding, one can increase q_L and u_L and generate strictly positive profit from low-risks while still respecting the incentive compatibility conditions. Now suppose $\lambda_H > 0$ and $\bar{\mu} > 0$. Then we must have $\varphi > 0$. In this case, (IC_H) reads $\sigma = 0$. We have therefore shown: $q_L = 0 \Rightarrow u_L = 0$.

Suppose $u_L = 0$. Then (**??**) reads $-\lambda_H + \varphi = 0$. This means we must either have $\lambda_H = 0$ and $\varphi = 0$ *or* $\lambda_H > 0$ and $\varphi > 0$. By the same arguments as before, we $\lambda_H = 0$ and $\varphi = 0$ cannot be a solution. Thus, we must have $\lambda_H > 0$ and $\varphi > 0$. In this case (IC_H) reads $\sigma = 0$. Therefore, we have shown $u_L \Rightarrow q_L = 0$.

Together the two preceding sets of arguments imply that a solution to the present sub-problem has $u_L = 0 \iff q_L = 0$. Moreover, a solution with $(u_L, q_L) = (0, 0)$ must also be such that $\lambda_H > 0$, so we have $u_H = \sigma = 0$. Therefore, we have shown that if $u_L = 0 \iff q_L = 0$ then $\sigma = 0$. Finally, note that since $u_H = \sigma$, if $\sigma = 0$, then IC_H implies u_L and $\Delta(q_L)$ are equal to zero. □

Lemma 10. *There exists a unique value of $\sigma > 0$ such that the boundary solution to the monopoly problem on $\mathcal{C}_2$ has $\lambda_H = 0$ and $q_L = q^*$ at $\sigma = \sigma^*$.*

Proof. From the previous lemma, we know that $q_L = 0$ and $\lambda_H > 0$ at $\sigma = 0$. We can therefore focus on $\sigma > 0$. Suppose $\lambda_H = 0$. We also know from the previous lemma that we must have $\varphi = 0$ for $\sigma > 0$. This, in turn, means we have $q_L = q^*$. Substituting into the IC constraints, we have $\Delta(q^*) \geq \sigma - u_L \geq \Delta(q^*)$. These conditions are satisfied if and only if $u_L = \sigma - \Delta(q^*)$. Because $\Delta(q^*)$ is constant, the remainder of our proof thus consists of showing that there exists a σ which solves the F.O.C. with respect to u_L at $u_L = \sigma - \Delta(q^*)$.

$$\frac{\partial G\Big(\sigma - \Delta(q^*); \sigma\Big)}{\partial u}\Big[s_L + \Delta(q^*) - \sigma\Big] - G\Big(\sigma - \Delta(q^*); \sigma\Big) = 0$$

Clearly when $\sigma \leq \Delta(q^*)$, the left-hand side of the F.O.C. is strictly positive. This means we can focus on the term in large parentheses in the following rearranged F.O.C.

$$\frac{G\Big(\sigma - \Delta(q^*); \sigma\Big)}{\sigma - \Delta(q^*)}\left(\eta\Big(\sigma - \Delta(q^*); \sigma\Big)\Big[s_L + \Delta(q^*) - \sigma\Big] + \Delta(q^*) - \sigma\right) = 0$$

Let $\Sigma(\sigma) \equiv \eta(\sigma - \Delta(q^*); \sigma)[s_L + \Delta(q^*) - \sigma] + \Delta(q^*) - \sigma$ and note that $\Sigma(\sigma)$ is continuous (and differentiable) for $\sigma \geq \Delta(q^*)$. Differentiating with respect to σ we have

$$\Sigma'(\sigma) = \mathcal{E}_u^\eta \cdot \frac{\Delta(q^*)}{\sigma} - \frac{\Delta(q^*)}{\sigma} - \eta(\sigma - \Delta(q^*);\sigma) < 0$$

Furthermore, at $\sigma = \Delta(q^*)$, we have $\Sigma(\Delta(q^*)) = \eta(0;\Delta(q^*))s_L^* > 0$ while $\Sigma(\sigma) \to -\infty$ as σ increases without bound. Therefore, there exists a unique σ, which we denote by σ^* at which $\lambda_H = 0$ and $q_L = q^*$ at the solution to the problem. □

Furthermore, we know that σ^* is implicitly defined by

$$\sigma - \Delta(q^*) = \frac{\eta\Big(\sigma - \Delta(q^*);\sigma\Big)}{1 + \eta\Big(\sigma - \Delta(q^*);\sigma\Big)} s_L^*$$

Sufficiency for $\sigma < \sigma^*$

Since $\lambda_L = 0$ any incentive compatible allocation has $u_H = u_L + \Delta(q_L)$. This means we can express any incentive compatible q_L as a function of u_H and u_L. In turn, this allows us to reformulate the problem as choosing $u \in [0,\sigma]$ to maximise

$$f_H[s_H^* - \sigma] + f_L G(u;\sigma)\Big[s\Big(q_L(\sigma,u);\theta_L\Big) - u\Big]$$

The objective is concave on $[0,\sigma]$ if and only if

$$\frac{\partial^2 G(u;\sigma)}{\partial(u)^2}\Big[s\Big(q_L(\sigma,u);\theta_L\Big) - u\Big] - 2\frac{\partial G(u;\sigma)}{\partial u} - 2\frac{\partial G(u;\sigma)}{\partial u}\Phi\Big(q_L(\sigma,u)\Big) + G(u;\sigma)\frac{\Phi'\Big(q_L(\sigma,u)\Big)}{\Delta'\Big(q_L(\sigma,u)\Big)} \leq 0 \tag{2.32}$$

Using Assumption **??**

$$
\begin{aligned}
&\frac{\partial^2 G(u;\sigma)}{\partial(u)^2}\Big[s\Big(q_L(\sigma,u);\theta_L\Big)-u\Big]-2\frac{\partial G(u;\sigma)}{\partial u}-2\frac{\partial G(u;\sigma)}{\partial u}\Phi\Big(q_L(\sigma,u)\Big)+G(u;\sigma)\frac{\Phi'\Big(q_L(\sigma,u)\Big)}{\Delta'\Big(q_L(\sigma,u)\Big)}\\
&\leq \frac{\partial G(u;\sigma)}{\partial u}\left(\frac{\frac{\partial^2 G(u;\sigma)}{\partial(u)^2}}{\frac{\partial G(u;\sigma)}{\partial u}}\Big[s\Big(q_L(\sigma,u);\theta_L\Big)-u\Big]-2-2\Phi\Big(q_L(\sigma,u)\Big)\right)\\
&\leq \frac{\partial G(u;\sigma)}{\partial u}\left(\frac{1}{G(u;\sigma)}\frac{\partial G(u;\sigma)}{\partial u}\Big[s\Big(q_L(\sigma,u);\theta_L\Big)-u\Big]-1-\Phi\Big(q_L(\sigma,u)\Big)-1-\Phi\Big(q_L(\sigma,u)\Big)\right)\\
&\leq \frac{\partial G(u;\sigma)}{\partial u}\left(-1-\Phi\Big(q_L(\sigma,u)\Big)\right)<0
\end{aligned}
$$

so the global concavity condition is satisfied.

Sufficiency for $\sigma \geq \sigma^*$

Because $\lambda_L = 0$ at any solution to the sub-problem, and because q_L is strictly increasing in σ whenever $\lambda_H > 0$ and $q_L < q^*$, we have that $q_L = q^*$ for all $\sigma \geq \sigma^*$.

Summary

- At $\sigma = 0$, the solution to the boundary problem is $\bar{C}_2 = \{(\sigma, q^*), (0,0)\}$.
- For $\sigma \in (0,\sigma^*)$, the solution to the boundary problem is $\bar{C}_2 = \{(\sigma, q^*), (\sigma-\Delta(q_L), q_L)\}$ with q_L given by

$$
\frac{G\Big(\sigma-\Delta(q_L);\sigma\Big)}{\sigma-\Delta(q_L)}\left(\eta\Big(\sigma-\Delta(q_L);\sigma\Big)\Big[s(q_L;\theta_L)+\Delta(q_L)-\sigma\Big]-[\sigma-\Delta(q^*)]\Big[1+\Phi(q_L)\Big]\right)=0
$$

- For $\sigma \geq \sigma^*$, the solution to the boundary problem is $\bar{C}_2 = \{(\sigma, q^*), (\sigma-\Delta(q^*), q^*)\}$.

Comparative statics

Comparative statics of u_L Suppose $\sigma \in (0,\sigma^*)$ so $\lambda_H > 0$. Substituting out q_L using $u_L = \sigma - \Delta(q_L)$, the remaining optimality condition reads

$$
0 = f_L G'\Big(\frac{u_L}{\sigma}\Big)\frac{1}{\sigma}\Big[s\Big(q_L(\sigma,u);\theta_L\Big)-u_L\Big]-f_L G\Big(\frac{u_L}{\sigma}\Big)-f_L G\Big(\frac{u_L}{\sigma}\Big)\Phi\Big(q_L(\sigma,u)\Big) \tag{2.33}
$$

The second partial derivative with respect to σ is

$$\frac{\partial^2 \pi}{\partial \sigma \partial u_L} = -\frac{u_L}{\sigma^2} f_L G''\Big(\frac{u_L}{\sigma}\Big)\frac{1}{\sigma}\Big[s\Big(q_L(\sigma,u);\theta_L\Big) - u_L\Big] - \frac{1}{\sigma} f_L G'\Big(\frac{u_L}{\sigma}\Big)\frac{1}{\sigma}\Big[s\Big(q_L(\sigma,u_L);\theta_L\Big) - u_L\Big]$$

$$+ \frac{u_L}{\sigma^2} f_L G'\Big(\frac{u_L}{\sigma}\Big) + \frac{u_L}{\sigma^2} f_L G'\Big(\frac{u_L}{\sigma}\Big)\Phi\Big(q_L(\sigma,u_L)\Big) + G'\Big(\frac{u_L}{\sigma}\Big)\frac{1}{\sigma}\Phi\Big(q_L(\sigma,u_L)\Big) - G\Big(\frac{u_L}{\sigma}\Big)\frac{\Phi'\Big(q_L(\sigma,u_L)\Big)}{\Delta'\Big(q_L(\sigma,u_L)\Big)}$$

$$= -f_L G\Big(\frac{u_L}{\sigma}\Big)\frac{1}{\sigma}\Big[1+\Phi\Big(q_L(\sigma,u_L)\Big)\Big]\Big(1 + \frac{u_L}{\sigma}\Big[\frac{G''(u_L/\sigma)}{G'(u_L/\sigma)} - \frac{G'(u_L/\sigma)}{G(u_L/\sigma)}\Big]\Big)$$

$$+ G\Big(\frac{u_L}{\sigma}\Big)\Bigg[\frac{G'(u_L/\sigma)}{G(u_L/\sigma)}\frac{1}{\sigma}\Phi\Big(q_L(\sigma,u_L)\Big) - \frac{\Phi'\Big(q_L(\sigma,u_L)\Big)}{\Delta'\Big(q_L(\sigma,u_L)\Big)}\Bigg]$$

The second term is strictly positive while the first term is strictly negative whenever $\eta(u;\sigma)$ is increasing in u. Thus, the sign of $\frac{du_L}{d\sigma}$ is in general indeterminate.

However, we have

$$\frac{du_L}{d\sigma} - 1 = \frac{G(u_L/\sigma)\frac{1}{\sigma}[1+\Phi(q_L)] + G'(u_L/\sigma)\frac{1}{\sigma} - G(u_L/\sigma)\frac{1}{\sigma}[1+\Phi(q_L)][1-u_L/\sigma][G''/G' - G'/G]}{G(u_L/\sigma)\frac{1}{\sigma}[1+\Phi(q_L)][G''/G' - G'/G] - G'(u_L/\sigma)\frac{1}{\sigma}[1+\Phi(q_L)] + G(u_L/\sigma)\frac{\Phi'}{\Delta'}}$$

By log-concavity of G, the denominator is strictly positive while the denominator is strictly negative. Hence $\frac{du_L}{d\sigma} < 1$.

Comparative statics of q_L Note that the binding IC_H condition implies $\Delta'(q_L)\frac{dq_L}{d\sigma} = 1 - \frac{du_L}{d\sigma}$. This means $\Delta'(q_L)\frac{dq_L}{d\sigma} > 0$.

Comparative statics of $G(u_L/\sigma)$ We thus have the following comparative statics of low-risk take-up

$$\frac{dG(u_L/\sigma)}{d\sigma} = G'\Big(\frac{u_L}{\sigma}\Big)\frac{1}{\sigma}\Bigg[\frac{\sigma}{u_L}\frac{du_L}{d\sigma} - 1\Bigg]\frac{u_L}{\sigma} \gtrless 0 \iff \frac{\sigma}{u_L}\frac{du_L}{d\sigma} \gtrless 1 \tag{2.34}$$

C.4 Interior solution on $\mathcal{C}_3$

We continue with our earlier approach of reformulating the monopoly problem on subspace $\mathcal{C}_3$ as the following simultaneous sub-problems

$$\max_{U} \quad f_H G(U/\sigma)[s_H^* - U] + f_L \pi(U) \qquad (\mathcal{P}_3^{outer})$$

$$where \quad \pi(U) \equiv \max_{(u,q)\in[0,\sigma]\times\mathbb{R}_+} \quad G(u/\sigma)[s(q;\theta_L) - u] \quad s.t. \quad u + \Delta(q) \leq U \qquad (\mathcal{P}^{inner})$$

Since the $\mathcal{P}^{inner}$ is identical to the case for $\mathcal{C}_2$, we focus immediately on $\mathcal{P}_3^{outer}$

Outer Problem

A solution to Problem $\mathcal{P}_3^{outer}$ must satisfy the F.O.C.

$$f_H G'\Big(\frac{U}{\sigma}\Big)\frac{1}{\sigma}[s_H^* - U] - f_H G\Big(\frac{U}{\sigma}\Big) + f_L \pi'(U) = 0 \quad where \quad \pi'(U) = G(u_L/\sigma)\Phi(q_L(U,u_L)) \tag{2.35}$$

Sufficiency Differentiating with respect to U gives

$$\begin{aligned}
&f_H G''\Big(\frac{U}{\sigma}\Big)\frac{1}{\sigma^2}[s_H^* - U] - 2f_H G'\Big(\frac{U}{\sigma}\Big)\frac{1}{\sigma} + \pi''(U)\\
&= f_H G''\Big(\frac{U}{\sigma}\Big)\frac{1}{\sigma^2}[s_H^* - U] - 2f_H G'\Big(\frac{U}{\sigma}\Big)\frac{1}{\sigma} + f_L G(u_L/\sigma)\frac{\Phi'(q_L(U,u_L))}{\Delta'(q_L(U,u_L))}\\
&+ f_L G(u_L/\sigma)\frac{du_L}{dU}\left(\frac{G'(u_L/\sigma)}{G(u_L/\sigma)}\frac{1}{\sigma}\Phi(q_L(U,u_L)) - \frac{\Phi'(q_L(U,u_L))}{\Delta'(q_L(U,u_L))}\right)
\end{aligned}$$

From our analysis of $\mathcal{P}^{inner}$, we have

$$\frac{du_L}{dU} = -\frac{G'(u_L/\sigma)\frac{1}{\sigma}\Phi(q_L) - G(u_L/\sigma)\frac{\Phi'(q_L)}{\Delta'(q_L)}}{G''(u_L/\sigma)\frac{1}{\sigma^2}[s_L - u_L] - 2G'(u_L/\sigma)\frac{1}{\sigma} - 2G'(u_L/\sigma)\frac{1}{\sigma}\Phi(q_L) + G(u_L/\sigma)\frac{\Phi'(q_L)}{\Delta'(q_L)}} \tag{2.36}$$

Substitution and algebra yields the following necessary and sufficient condition for concavity of the objective function

$$\begin{aligned}
&\frac{f_H}{f_L}G'(u_H/\sigma)\frac{1}{\sigma}f_L G'(u_L/\sigma)\frac{1}{\sigma}\left(\frac{G''(u_H/\sigma)}{G'(u_H/\sigma)}\frac{1}{\sigma}[s_H^*-u_H]-2\right)\left(\frac{G''(u_L/\sigma)}{G'(u_L/\sigma)}\frac{1}{\sigma}[s(q_L;\theta_L)-u_L]-2\right)\\
&-2\frac{f_H}{f_L}G'(u_H/\sigma)\frac{1}{\sigma}f_L G'(u_L/\sigma)\frac{1}{\sigma}\Phi(q_L)\left(\frac{G''(u_H/\sigma)}{G'(u_H/\sigma)}\frac{1}{\sigma}[s_H^*-u_H]-2\right)\\
&+\frac{f_H}{f_L}G'(u_H/\sigma)\frac{1}{\sigma}G(u_L/\sigma)\frac{\Phi'(q_L)}{\Delta'(q_L)}\left(\frac{G''(u_H/\sigma)}{G'(u_H/\sigma)}\frac{1}{\sigma}[s_H^*-u_H]-2\right)\\
&+G'(u_L/\sigma)\frac{1}{\sigma}G(u_L/\sigma)\frac{\Phi'(q_L)}{\Delta'(q_L)}\left(\frac{G''(u_L/\sigma)}{G'(u_L/\sigma)}\frac{1}{\sigma}[s(q_L;\theta_L)-u_L]-2\right)-\left(G'(u_L/\sigma)\frac{1}{\sigma}\Phi(q_L)\right)^2\geq 0
\end{aligned}$$

Note that every term except that last is positive. Furthermore, under our assumptions on G, $G'(u_L/\sigma)\frac{1}{\sigma}\Phi(q_L)$ is bounded above by $G'(1)\Phi(0)/\sigma$. This implies that for sufficiently large f_H/f_L, the concavity condition is satisfied.

Lemma 11. *There exists a unique $\hat{\sigma}_H$ such that $U=\hat{\sigma}_H$ solves F.O.C. (2.35). Furthermore, the corresponding solution to Problem $\mathcal{P}^{inner}$ is $(u_{L,3},q_{L,3})=(u_{L,2},q_{L,2})$.*

Proof. Suppose $U=\sigma$. Then F.O.C. (2.35) reads

$$\Gamma(\sigma)\equiv f_H G'(1)\frac{1}{\sigma}[s_H^*-\sigma]-f_H+\pi'(\sigma)=0 \tag{2.37}$$

At $\sigma=s_H^*$, $\Gamma(\sigma)=-f_H+\pi'(\sigma)<0$. The inequality obtains from the fact that $-f_H+\pi'(U)<0$ for all $U\geq 0$ when $\sigma>\hat{\sigma}_L$ and from noting that $\hat{\sigma}_L<s_H^*$. On the other hand, taking $\sigma<\hat{\sigma}_L$, we have $-f_H+\pi'(U)>0$ for all U so when σ is sufficiently small, we have $\Gamma(\sigma)>0$. Finally, differentiating with respect to σ,

$$\Gamma'(\sigma)=-f_H G'(1)\frac{1}{\sigma}+\left.\frac{d\pi'(U)}{d\sigma}\right|_{U=\sigma}<0$$

since $d\pi'(U)/d\sigma<0$ for any given U. Hence, there exists a unique $\hat{\sigma}_H>\hat{\sigma}_L$ such that $U=\hat{\sigma}_H$ solves (2.35). By construction, corresponding the solution to $\mathcal{P}^{inner}$ coincides with the boundary solution on $\mathcal{C}_2$. □

Note also that since $\Gamma'(\sigma)<0$ we also know that F.O.C. (2.35) reads $f_H G'(1)\frac{1}{\sigma}[s_H^*-\sigma]-f_H+\pi'(\sigma)>0$ for $\sigma<\hat{\sigma}_H$. This means for $\sigma<\hat{\sigma}_H$ the solution on $\mathcal{C}_3$ is on the boundary where $u_H=\sigma$. In turn, applying previous arguments, we know this allocation coincides with the boundary solution on $\mathcal{C}_2$.

D Existence of a solution to $\mathcal{P}^m$

Lemma 12. *Assume Assumptions 1, 2, 3 hold and $G(z)$ is log-concave and satisfies Assumption 4. For all $\sigma \geq 0$, a solution to the monopoly problem on each of the sub-spaces $\mathcal{C}_1$, $\mathcal{C}_2$ and $\mathcal{C}_3$ exists.*

Proof. Under the incentive compatibility conditions, the solution space can be partitioned in $\mathcal{C}_1$, $\mathcal{C}_2$ and $\mathcal{C}_3$. Because each of these sub-spaces is compact and because the monopolist's objective function is continuous on each of these sub-spaces, there exists a solution to the sub-problem for each $\mathcal{C}_r$, $r = 1, 2, 3$. Moreover we have shown that a unique solution to the sub-problem on $\mathcal{C}_1$ is at the boundary where $u_L = \sigma$ and the unique solution to the sub-problem on $\mathcal{C}_2$ must be at the boundary where $u_H = \sigma$. For $\mathcal{C}_3$, the solution can be on the interior of $\mathcal{C}_3$, at the boundary where $u_H = \sigma$, or the boundary where $u_L = 0$. □

The previous lemma argued that there is a solution on each of the sub-spaces for all $\sigma \geq 0$. The subsequent lemmas show where each of these candidates is the global optimum.

D.1 Threshold value $\bar{\sigma}_1$

Lemma 13. *Let $C_1 \equiv \{(u_{H,1}, q_{H,1}), (u_{L,1}, q_{L,1})\}$ denote the solution to the monopoly problem on $\mathcal{C}_1$ and let $C_2 \equiv \{(u_{H,2}, q_{H,2}), (u_{L,2}, q_{L,2})\}$ denote the solution to the monopoly problem on $\mathcal{C}_2$. For each $\frac{f_H}{f_L} \geq 0$, there exists a unique critical value, $\bar{\sigma}_1 \geq 0$, such that $\Pi(C_1) < \Pi(C_2)$ if and only if $\sigma > \bar{\sigma}_1$.*

Proof. When $\frac{f_H}{f_L} \geq \Phi(0)$, $C_1 = \{(\sigma, q^*), (\sigma, 0)\}$, while there is no interior solution on $\mathcal{C}_2$ for all $\sigma \geq 0$, so C_2 is the boundary solution with $u_{H,2} = \sigma$. Since $s(q_{L,2}; \theta_L) - u_{L,2} \geq 0$, we have

$$\Pi(C_1) - \Pi(C_2) = \left\{ f_H[s_H^* - \sigma] + f_L[-\sigma] \right\} - \left\{ f_H[s_H^* - \sigma] + f_L G(u_{L,2}; \sigma)\Big[s(q_{L,2}; \theta_L) - u_{L,2} \Big] \right\} < 0$$

This means the critical value when $\frac{f_H}{f_L} \geq \Phi(0)$ is $\bar{\sigma}_1 = 0$.

Consider $\frac{f_H}{f_L} < \Phi(0)$. Recall that we defined $\hat{\sigma}_L$ as the value given by $G'(1)[s(q_L^\infty; \theta_L) - \hat{\sigma}_L + \Delta(q_L^\infty)] - 1 - \Phi(q_L^\infty) = 0$. Furthermore, we argued that for $\sigma \leq \hat{\sigma}_L$, C_2 is such that $u_{L,2} = \sigma$. Since C_2 and C_1 coincide whenever $u_{L,2} = \sigma$, we have $\Pi(C_1) = \Pi(C_2)$ for $\sigma \leq \hat{\sigma}_L$.

For $\sigma > \hat{\sigma}_L$, we showed that the monopolist's total expected profit on $\mathcal{C}_2$ is strictly decreasing in u_H for all $u_H \geq \sigma$. Therefore, a solution to the monopoly problem on $\mathcal{C}_2$ must be on the boundary with $u_{H,2} = \sigma$. Let $\bar{C}_2$ denote this boundary solution. For any $\hat{C} \in \mathcal{C}_2$ we have $\Pi(\bar{C}_2) \geq \Pi(\hat{C})$, where the inequality is strict whenever $\hat{C}$ has $u_H > \sigma$. Since C_1 yields the same total expected profit as $\hat{C} = \{(u_H + \Delta(q_L^\infty), q^*), (\sigma, q_L^\infty)\}$, we therefore have $\Pi(\bar{C}_2) > \Pi(C_1)$ for all $\sigma > \hat{\sigma}_L$.

In summary, the unique critical value, $\bar{\sigma}_1 \geq 0$, such that $\Pi(C_1) < \Pi(C_2)$ if and only if $\sigma \geq \bar{\sigma}_1$ is given by $\bar{\sigma}_1 = 0$ for $f_H/f_L \geq \Phi(0)$ and $\bar{\sigma}_1 = \hat{\sigma}_L$ for $f_H/f_L < \Phi(0)$. □

Corollary 5. *The threshold value $\bar{\sigma}_1$ is weakly decreasing in $\frac{f_H}{f_L}$.*

Proof. For $f_H/f_L \geq \Phi(0)$, $\bar{\sigma}_1 = 0$ is invariant in f_H/f_L. Consider $f_H/f_L < \Phi(0)$. Using $f_H/f_L = \Phi(q_L^\infty)$ in the definition of $\hat{\sigma}_L$, we have

$$G'(1)[s(q_L^\infty;\theta_L) - \hat{\sigma}_L + \Delta(q_L^\infty)] - 1 - \frac{f_H}{f_L} = 0$$

Differentiating with respect to f_H/f_L, we have

$$G'(1)s_q(q_L^\infty;\theta_L)\frac{dq^\infty}{df_H/f_L}df_H/f_L + G'(1)\Delta'(q_L^\infty)\frac{dq^\infty}{df_H/f_L}df_H/f_L - G'(1)d\hat{\sigma}_L - df_H/f_L = 0$$

As we have shown above that $dq_L^\infty/df_H/f_L < 0$, we have

$$\frac{d\hat{\sigma}_L}{df_H/f_L} = [1 + \Phi(q_L^\infty)]\Delta'(q_L^\infty)\frac{dq_L^\infty}{df_H/f_L} - \frac{1}{G'(1)} < 0$$

We have thus shown that $\bar{\sigma}_1$ is weakly decreasing in $\frac{f_H}{f_L}$ for all $\frac{f_H}{f_L}$. □

D.2 Threshold value $\bar{\sigma}_2$

Lemma 14. *Let $C_2 \equiv \{(u_{H,2}, q_{H,2}), (u_{L,2}, q_{L,2})\}$ denote the solution to the monopoly problem on $\mathcal{C}_2$ and let $C_3 \equiv \{(u_{H,3}, q_{H,3}), (u_{L,3}, q_{L,3})\}$ denote the solution to the monopoly problem on $\mathcal{C}_3$. There exists a unique critical value, $\bar{\sigma}_2 \geq 0$, such that $\Pi(C_2) < \Pi(C_3)$ if and only if $\sigma > \bar{\sigma}_2$.*

Proof. First suppose that the objective is concave on $\mathcal{C}_3$. We have shown that a sufficient condition for concavity is to have sufficiently large f_H/f_L. Given that an interior solution on $\mathcal{C}_3$ exists, recall that we argued that for $\sigma \leq \hat{\sigma}_H$, C_3 is such that $u_{H,3} = \sigma$. Since C_3 and

C_2 coincide when $u_{H,3} = \sigma$, we have $\Pi(C_3) = \Pi(C_2)$ for $\sigma \leq \hat{\sigma}_H$. Conversely, for $\sigma > \hat{\sigma}_H$, we know that C_3 necessarily has $u_{H,3} < \sigma$. Therefore, $\Pi(C_3) > \Pi(C_2)$ for $\sigma > \hat{\sigma}_H$.

Now suppose that an interior solution on $\mathcal{C}_3$ does not exist. In this case the monopolist's profit is maximised either when $u_H = \sigma$ or when $u_L = 0$. Let the profit-maximising menu in the former case be denoted by $\bar{C}_2 = \{(\sigma, q^*), (u_{L,2}, q_{L,2})\}$ since we have shown that it corresponds to the boundary solution on $\mathcal{C}_2$. In the latter case, we can show that the profit-maximising menu is $(u_L, q_L) = (0,0)$ with $q_H = q^*$ and u_H given by $\eta(u_H;\sigma)[s_H^*/u_H - 1] - 1 = 0$. The difference in total expected profit from offering $\bar{C}_2$ versus $\bar{C}_3$ is

$$\Pi(\bar{C}_2) - \Pi(\bar{C}_3) = \Bigg\{ f_H[s_H^* - \sigma] + f_L G(u_{L,2};\sigma)\Big[s(q_{L,2};\theta_L) - u_{L,2}\Big]\Bigg\} - f_H G(u_{H,3};\sigma)[s_H^* - u_{H,3}]$$

Let $\Omega_2(\sigma) \equiv \Pi(\bar{C}_2) - \Pi(\bar{C}_3)$ and note that it is continuous in σ since the value function $\Pi(\bar{C}_2)$ and $\Pi(\bar{C}_3)$ are continuous under our baseline assumptions. From our analysis of the benchmark problem, for $\sigma \leq \sigma_H$, we have $u_{H,3} = \sigma$. This implies $\Omega_2(\sigma) > 0$ for $\sigma \in (0, \sigma_H]$. Next, since $s(q_{L,2};\theta_L) - u_{L,2}$ takes its maximum value at $s_L^* + \Delta(q^*) - \sigma$, at $\sigma = s_L^* + \Delta(q^*)$ the term in curly braces is equal to $-f_H[c(q^*;\theta_H) - c(q^*;\theta_L)] < 0$. Thus, $\Omega_2(\sigma) < 0$ for $\sigma \geq s_L^* + \Delta(q^*)$.

Finally, differentiating with respect to σ using the envelope condition yields

$$\begin{aligned}
\Omega_2'(\sigma) &= \Bigg\{ - f_H - f_L G(u_{L,2};\sigma)\frac{u_L}{\sigma}\Bigg[\frac{\eta(u_{L,2};\sigma)}{u_{L,2}}[s(q_{L,2};\theta_L) - u_{L,2}]\Bigg]\Bigg\} - f_H\Bigg\{ - G(u_{H,3};\sigma)\frac{u_{H,3}}{\sigma}\frac{\eta(u_{H,3};\sigma)}{u_{H,3}}[s_H^* - u_{H,3}]\Bigg\} \\
&= \Bigg\{ - f_H - f_L G(u_{L,2};\sigma)\frac{u_L}{\sigma}\Bigg[1 + \frac{\lambda_{H,2}}{f_L G(u_{L,2};\theta_L)}\Bigg]\Bigg\} - f_H\Bigg\{ - G(u_{H,3};\sigma)\frac{u_{H,3}}{\sigma}\Bigg\} \\
&= - f_L G(u_{L,2};\sigma)\frac{u_L}{\sigma}\Bigg[1 + \frac{\lambda_{H,2}}{f_L G(u_{L,2};\theta_L)}\Bigg] - f_H\Bigg[1 - G(u_{H,3};\sigma)\frac{u_{H,3}}{\sigma}\Bigg] < 0
\end{aligned}$$

Therefore, there exists a unique critical value of σ, call this $\tilde{\sigma}_H$, such that $\Pi(\bar{C}_2) < \Pi(\bar{C}_3) \iff \sigma > \tilde{\sigma}_H$.

In summary, the unique critical value, $\bar{\sigma}_2$, such that $\Pi(C_2) < \Pi(C_3)$ if and only if $\sigma > \bar{\sigma}_2$ is given by $\bar{\sigma}_2 = \hat{\sigma}_H$ when an interior solution to the monopoly problem on $\mathcal{C}_3$ exists and $\bar{\sigma}_2 = \tilde{\sigma}_H$ otherwise. $\square$

By the preceding lemmas, for $\sigma \leq \bar{\sigma}_1 < \bar{\sigma}_2$, we have $\Pi(C_1) \geq \Pi(C_2) > \Pi(C_3)$ so C_1 is globally optimal. For $\bar{\sigma}_1 < \sigma \leq \bar{\sigma}_2$, we have $\Pi(C_1) < \Pi(C_2)$ and $\Pi(C_2) \geq \Pi(C_3)$ so C_2 is globally optimal. For $\sigma > \bar{\sigma}_2$, $\Pi(C_3) > \Pi(C_2) > \Pi(C_1)$ so C_3 is globally optimal.

E Comparative statics

E.1 Comparative statics of C_3

Since $u_H - u_L = \Delta(q_L)$ at any solution on $\mathcal{C}_3$, we can substitute out q_L in the insurer's problem. The resulting F.O.C.s for this resulting problem, with choice variables (u_H, u_L), are

$$f_H G\Big(\frac{u_H}{\sigma}\Big)\eta(u_H;\sigma)\frac{1}{u_H}[s_H^* - u_H] - f_H G\Big(\frac{u_H}{\sigma}\Big) + f_L G\Big(\frac{u_L}{\sigma}\Big)\Phi\Big(q_L(u_H,u_L)\Big) = 0 \tag{2.38}$$

$$f_L G\Big(\frac{u_L}{\sigma}\Big)\eta(u_L;\sigma)\frac{1}{u_L}\Big[s\Big(q_L(u_H,u_L;\theta_L)\Big) - u_L\Big] - f_L G\Big(\frac{u_L}{\sigma}\Big) - f_L G\Big(\frac{u_L}{\sigma}\Big)\Phi\Big(q_L(u_H,u_L)\Big) = 0 \tag{2.39}$$

The cross-partial derivatives of the objective with respect to σ are

$$\begin{aligned}
\frac{\partial^2\Pi}{\partial\sigma\partial u_H} =& -\frac{u_H}{\sigma^2} f_H G'\Big(\frac{u_H}{\sigma}\Big)\eta(u_H;\sigma)\frac{1}{u_H}[s_H^* - u_H] + f_H G\Big(\frac{u_H}{\sigma}\Big)\frac{\partial\eta(u_H;\sigma)}{\partial\sigma}\frac{1}{u_H}[s_H^* - u_H] \\
&+ \frac{u_H}{\sigma^2} f_H G'\Big(\frac{u_H}{\sigma}\Big) - \frac{u_L}{\sigma^2} f_L G'\Big(\frac{u_L}{\sigma}\Big)\Phi\Big(q_L(u_H,u_L)\Big) \\
=& f_H G\Big(\frac{u_H}{\sigma}\Big)\frac{\partial\eta(u_H;\sigma)}{\partial\sigma}\frac{1}{\eta(u_H;\sigma)}\Big[1 - \frac{f_L}{f_H}\frac{G(u_L/\sigma)}{G(u_H/\sigma)}\Phi\Big(q_L(u_H,u_L)\Big)\Big] \\
&+ f_L G\Big(\frac{u_L}{\sigma}\Big)\frac{1}{\sigma}\Phi\Big(q_L(u_H,u_L)\Big)\Big[\eta(u_H;\sigma) - \eta(u_L;\sigma)\Big] \\
\frac{\partial^2\Pi}{\partial\sigma\partial u_L} =& -\frac{u_L}{\sigma^2} f_L G'\Big(\frac{u_L}{\sigma}\Big)\eta(u_L;\sigma)\frac{1}{u_L}\Big[s\Big(q_L(u_H,u_L;\theta_L)\Big) - u_L\Big] + f_L G\Big(\frac{u_L}{\sigma}\Big)\frac{\partial\eta(u_L;\sigma)}{\partial\sigma}\frac{1}{u_L}\Big[s\Big(q_L(u_H,u_L;\theta_L)\Big) - u_L\Big] \\
&+ \frac{u_L}{\sigma^2} f_L G'\Big(\frac{u_L}{\sigma}\Big) + \frac{u_L}{\sigma^2} f_L G'\Big(\frac{u_L}{\sigma}\Big)\Phi\Big(q_L(u_H,u_L)\Big) \\
=& f_L G\Big(\frac{u_L}{\sigma}\Big)\frac{\partial\eta(u_L;\sigma)}{\partial\sigma}\frac{1}{\eta(u_L;\sigma)}\Big[1 + \Phi\Big(q_L(u_H,u_L)\Big)\Big]
\end{aligned}$$

Note that

$$\frac{\partial\eta(u;\sigma)}{\partial\sigma} = -\frac{u}{\sigma}\left(\frac{u}{\sigma}\Big[\frac{d}{dz}\frac{G'(z)}{G(z)}\Big]\frac{1}{\sigma} + \frac{G'(u/\sigma)}{G(u/\sigma)}\frac{1}{\sigma}\right) = -\frac{u}{\sigma}\frac{\partial\eta(u;\sigma)}{\partial u}$$

So $\eta(u;\sigma)$ is decreasing in σ since it is increasing in u and

$$\begin{aligned}
\frac{\partial^2\Pi}{\partial\sigma\partial u_H} =& -\frac{u_H}{\sigma}f_H G\Big(\frac{u_H}{\sigma}\Big)\frac{\partial\eta(u_H;\sigma)}{\partial u}\frac{1}{\eta(u_H;\sigma)}\Big[1-\frac{f_L}{f_H}\frac{G(u_L/\sigma)}{G(u_H/\sigma)}\Phi\Big(q_L(u_H,u_L)\Big)\Big]\\
&+f_L G\Big(\frac{u_L}{\sigma}\Big)\frac{1}{\sigma}\Phi\Big(q_L(u_H,u_L)\Big)\Big[\eta(u_H;\sigma)-\eta(u_L;\sigma)\Big]\\
\frac{\partial^2\Pi}{\partial\sigma\partial u_L} =& -\frac{u_L}{\sigma}f_L G\Big(\frac{u_L}{\sigma}\Big)\frac{\partial\eta(u_L;\sigma)}{\partial u}\frac{1}{\eta(u_L;\sigma)}\Big[1+\Phi\Big(q_L(u_H,u_L)\Big)\Big]
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2\Pi}{\partial(u_H)^2} =& f_H G'\Big(\frac{u_H}{\sigma}\Big)\frac{1}{\sigma}\eta(u_H;\sigma)\frac{1}{u_H}[s_H^*-u_H]+f_H G\Big(\frac{u_H}{\sigma}\Big)\frac{\partial\eta(u_H;\sigma)}{\partial u_H}\frac{1}{u_H}[s_H^*-u_H]\\
&-f_H G\Big(\frac{u_H}{\sigma}\Big)\eta(u_H;\sigma)\frac{1}{u_H}\left(1+\frac{1}{u_H}[s_H^*-u_H]\right)-f_H G'\Big(\frac{u_H}{\sigma}\Big)\frac{1}{\sigma}+f_L G\Big(\frac{u_L}{\sigma}\Big)\frac{\Phi'(q_L)}{\Delta'(q_L)}\\
=& f_H G\Big(\frac{u_H}{\sigma}\Big)\frac{\partial\eta(u_H;\sigma)}{\partial u_H}\frac{1}{u_H}\Big[1-\frac{f_L}{f_H}\frac{G(u_L/\sigma)}{G(u_H/\sigma)}\Phi(q_L)\Big]-f_H G\Big(\frac{u_H}{\sigma}\Big)\frac{1}{u_H}\Big[1-\frac{f_L}{f_H}\frac{G(u_L/\sigma)}{G(u_H/\sigma)}\Phi(q_L)\Big]\\
&-f_H G\Big(\frac{u_H}{\sigma}\Big)\eta(u_H;\sigma)\frac{1}{u_H}-f_H G\Big(\frac{u_H}{\sigma}\Big)\eta(u_H;\sigma)\frac{1}{u_H}\frac{f_L}{f_H}\frac{G(u_L/\sigma)}{G(u_H/\sigma)}\Phi(q_L)+f_L G\Big(\frac{u_L}{\sigma}\Big)\frac{\Phi'(q_L)}{\Delta'(q_L)}\\
=& f_H G\Big(\frac{u_H}{\sigma}\Big)\frac{\partial\eta(u_H;\sigma)}{\partial u_H}\frac{1}{u_H}\Big[1-\frac{f_L}{f_H}\frac{G(u_L/\sigma)}{G(u_H/\sigma)}\Phi(q_L)\Big]-f_H G\Big(\frac{u_H}{\sigma}\Big)\frac{1}{u_H}\Big[1-\frac{f_L}{f_H}\frac{G(u_L/\sigma)}{G(u_H/\sigma)}\Phi(q_L)\Big]\\
&-f_H G\Big(\frac{u_H}{\sigma}\Big)\eta(u_H;\sigma)\frac{1}{u_H}\Big[1-\frac{f_L}{f_H}\frac{G(u_L/\sigma)}{G(u_H/\sigma)}\Phi(q_L)\Big]\\
&-2f_L G\Big(\frac{u_L}{\sigma}\Big)\eta(u_H;\sigma)\frac{1}{u_H}\Phi(q_L)+f_L G\Big(\frac{u_L}{\sigma}\Big)\frac{\Phi'(q_L)}{\Delta'(q_L)}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2\Pi}{\partial(u_L)^2} =& f_L G'\Big(\frac{u_L}{\sigma}\Big)\frac{1}{\sigma}\eta(u_L;\sigma)\frac{1}{u_L}\Big[s\Big(q_L(u_H,u_L;\theta_L)\Big)-u_L\Big]+f_L G\Big(\frac{u_L}{\sigma}\Big)\frac{\partial\eta(u_L;\sigma)}{\partial u}\frac{1}{u_L}\Big[s\Big(q_L(u_H,u_L;\theta_L)\Big)-u_L\Big]\\
&-f_L G\Big(\frac{u_L}{\sigma}\Big)\eta(u_L;\sigma)\frac{1}{u_L}\left(1+\Phi(q_L)+\frac{1}{u_L}\Big[s\Big(q_L(u_H,u_L;\theta_L)\Big)-u_L\Big]\right)\\
&-f_L G'\Big(\frac{u_L}{\sigma}\Big)\frac{1}{\sigma}-f_L G'\Big(\frac{u_L}{\sigma}\Big)\Phi\Big(q_L(u_H,u_L)\Big)-f_L G\Big(\frac{u_L}{\sigma}\Big)\frac{\Phi'(q_L)}{\Delta'(q_L)}\\
=& f_L G\Big(\frac{u_L}{\sigma}\Big)\frac{\partial\eta(u_L;\sigma)}{\partial u}\frac{1}{u_L}\Big[1+\Phi(q_L)\Big]-f_L G\Big(\frac{u_L}{\sigma}\Big)\frac{1}{u_L}\Big[1+\Phi(q_L)\Big]\\
&-f_L G\Big(\frac{u_L}{\sigma}\Big)\eta(u_L;\sigma)\frac{1}{u_L}\Big[1+\Phi(q_L)\Big]+f_L G\Big(\frac{u_L}{\sigma}\Big)\frac{\Phi'(q_L)}{\Delta'(q_L)}\\
\frac{\partial^2\Pi}{\partial u_H\partial u_L} =& f_L G\Big(\frac{u_L}{\sigma}\Big)\frac{1}{u_L}\eta(u_L;\sigma)\Phi\Big(q_L(u_H,u_L)\Big)-f_L G\Big(\frac{u_L}{\sigma}\Big)\frac{\Phi'(q_L)}{\Delta'(q_L)}
\end{aligned}$$

Using the IC_H relation $u_H - u_L = \Delta(q_L)$, we have $\Delta'(q_L)dq_{L,2}/d\sigma = du_{H,2}/d\sigma - du_{L,2}/d\sigma$. Therefore

$$
\begin{aligned}
\Delta'(q_L)\frac{dq_{L,2}}{d\sigma} \gtreqless 0 \iff & -\frac{\partial^2\Pi}{\partial\sigma\partial u_H}\frac{\partial^2\Pi}{\partial(u_L)^2} + \frac{\partial^2\Pi}{\partial u_H\partial u_L}\frac{\partial^2\Pi}{\partial\sigma\partial u_L} - \left(-\frac{\partial^2\Pi}{\partial(u_H)^2}\frac{\partial^2\Pi}{\partial\sigma\partial u_L} + \frac{\partial^2\Pi}{\partial\sigma\partial u_H}\frac{\partial^2\Pi}{\partial u_H\partial u_L} \right) \gtreqless 0 \\
\iff & -\frac{\partial^2\Pi}{\partial\sigma\partial u_H}\frac{\partial^2\Pi}{\partial(u_L)^2} + \frac{\partial^2\Pi}{\partial u_H\partial u_L}\frac{\partial^2\Pi}{\partial\sigma\partial u_L} + \frac{\partial^2\Pi}{\partial(u_H)^2}\frac{\partial^2\Pi}{\partial\sigma\partial u_L} - \frac{\partial^2\Pi}{\partial\sigma\partial u_H}\frac{\partial^2\Pi}{\partial u_H\partial u_L} \gtreqless 0 \\
\iff & -\frac{\partial^2\Pi}{\partial\sigma\partial u_H}\left(\frac{\partial^2\Pi}{\partial(u_L)^2} + \frac{\partial^2\Pi}{\partial u_H\partial u_L}\right) + \left(\frac{\partial^2\Pi}{\partial(u_H)^2} + \frac{\partial^2\Pi}{\partial u_H\partial u_L}\right)\frac{\partial^2\Pi}{\partial\sigma\partial u_L} \gtreqless 0
\end{aligned}
$$

where

$$
\begin{aligned}
\frac{\partial^2\Pi}{\partial(u_L)^2} + \frac{\partial^2\Pi}{\partial u_H\partial u_L} =& \frac{\partial\eta(u_L;\sigma)}{\partial u} f_L G\Big(\frac{u_L}{\sigma}\Big)\frac{1}{u_L}\Big[1+\Phi(q_L)\Big] - f_L G\Big(\frac{u_L}{\sigma}\Big)\frac{1}{u_L}\Big[1+\Phi(q_L)\Big] - \eta(u_L;\sigma) f_L G\Big(\frac{u_L}{\sigma}\Big)\frac{1}{u_L}\Big[1+\Phi(q \\
& + f_L G\Big(\frac{u_L}{\sigma}\Big)\frac{1}{u_L}\eta(u_L;\sigma)\Phi(q_L)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2\Pi}{\partial(u_H)^2} + \frac{\partial^2\Pi}{\partial u_H\partial u_L} =& \frac{\partial\eta(u_H;\sigma)}{\partial u_H} f_H G\Big(\frac{u_H}{\sigma}\Big)\frac{1}{u_H}\left[1 - \frac{f_L}{f_H}\frac{G(u_L/\sigma)}{G(u_H/\sigma)}\Phi(q_L)\right] - f_H G\Big(\frac{u_H}{\sigma}\Big)\frac{1}{u_H}\left[1 - \frac{f_L}{f_H}\frac{G(u_L/\sigma)}{G(u_H/\sigma)}\Phi(q_L)\right] \\
& - \eta(u_H;\sigma) f_H G\Big(\frac{u_H}{\sigma}\Big)\frac{1}{u_H}\left[1 - \frac{f_L}{f_H}\frac{G(u_L/\sigma)}{G(u_H/\sigma)}\Phi(q_L)\right] \\
& - 2f_L G\Big(\frac{u_L}{\sigma}\Big)\eta(u_H;\sigma)\frac{1}{u_H}\Phi(q_L) + f_L G\Big(\frac{u_L}{\sigma}\Big)\frac{1}{u_L}\eta(u_L;\sigma)\Phi(q_L)
\end{aligned}
$$

So

$$
\begin{aligned}
& -\frac{\partial^2\Pi}{\partial\sigma\partial u_H}\left(\frac{\partial^2\Pi}{\partial(u_L)^2} + \frac{\partial^2\Pi}{\partial u_H\partial u_L}\right) + \left(\frac{\partial^2\Pi}{\partial(u_H)^2} + \frac{\partial^2\Pi}{\partial u_H\partial u_L}\right)\frac{\partial^2\Pi}{\partial\sigma\partial u_L} \\
& = \left\{ \frac{u_H}{\sigma} f_H G\Big(\frac{u_H}{\sigma}\Big)\frac{\partial\eta(u_H;\sigma)}{\partial u}\frac{1}{\eta(u_H;\sigma)}\left[1 - \frac{f_L}{f_H}\frac{G(u_L/\sigma)}{G(u_H/\sigma)}\Phi(q_L)\right] - f_L G\Big(\frac{u_L}{\sigma}\Big)\frac{1}{\sigma}\Phi(q_L)\Big[\eta(u_H;\sigma) - \eta(u_L;\sigma)\Big] \right\} \\
& \times \left\{ \left(\frac{\partial\eta(u_L;\sigma)}{\partial u} - 1 - \eta(u_L;\sigma)\right) f_L G\Big(\frac{u_L}{\sigma}\Big)\frac{1}{u_L}\Big[1+\Phi(q_L)\Big] + f_L G\Big(\frac{u_L}{\sigma}\Big)\frac{1}{u_L}\eta(u_L;\sigma)\Phi(q_L) \right\} \\
& - \frac{u_L}{\sigma} f_L G\Big(\frac{u_L}{\sigma}\Big)\frac{\partial\eta(u_L;\sigma)}{\partial u}\frac{1}{\eta(u_L;\sigma)}\Big[1+\Phi(q_L)\Big] \\
& \times \left\{ \left(\frac{\partial\eta(u_H;\sigma)}{\partial u_H} - 1 - \eta(u_H;\sigma)\right) f_H G\Big(\frac{u_H}{\sigma}\Big)\frac{1}{u_H}\left[1 - \frac{f_L}{f_H}\frac{G(u_L/\sigma)}{G(u_H/\sigma)}\Phi(q_L)\right] \right. \\
& \left. - 2f_L G\Big(\frac{u_L}{\sigma}\Big)\eta(u_H;\sigma)\frac{1}{u_H}\Phi(q_L) + f_L G\Big(\frac{u_L}{\sigma}\Big)\frac{1}{u_L}\eta(u_L;\sigma)\Phi(q_L) \right\}
\end{aligned}
$$

$$
\begin{aligned}
& -\frac{\partial^2\Pi}{\partial\sigma\partial u_H}\frac{\partial^2\Pi}{\partial(u_L)^2}+\frac{\partial^2\Pi}{\partial u_H\partial u_L}\frac{\partial^2\Pi}{\partial\sigma\partial u_L}+\frac{\partial^2\Pi}{\partial(u_H)^2}\frac{\partial^2\Pi}{\partial\sigma\partial u_L}-\frac{\partial^2\Pi}{\partial\sigma\partial u_H}\frac{\partial^2\Pi}{\partial u_H\partial u_L} \\
=& -\frac{\partial^2\Pi}{\partial\sigma\partial u_H}\left(\frac{\partial^2\Pi}{\partial(u_L)^2}+\frac{\partial^2\Pi}{\partial u_H\partial u_L}\right)+\left(\frac{\partial^2\Pi}{\partial(u_H)^2}+\frac{\partial^2\Pi}{\partial u_H\partial u_L}\right)\frac{\partial^2\Pi}{\partial\sigma\partial u_L} \\
=& \left(\frac{u_H}{\sigma}f_H G_H\frac{\partial\eta(u_H;\sigma)}{\partial u}\frac{1}{\eta(u_H;\sigma)}[1-\frac{f_L}{f_H}\frac{G_L}{G_H}\Phi]-f_L G_L\frac{1}{\sigma}\Phi\Big[\eta(u_H;\sigma)-\eta(u_L;\sigma)\Big]\right) \\
& \left(f_L G_L\frac{\partial\eta(u_L;\sigma)}{\partial u}\frac{1}{u_L}[1+\Phi]-f_L G_L\frac{1}{u_L}-f_L G_L\eta(u_L;\sigma)\frac{1}{u_L}[1+\Phi]\right) \\
& +\left(f_H G_H\frac{\partial\eta(u_H;\sigma)}{\partial u_H}\frac{1}{u_H}\left[1-\frac{f_L}{f_H}\frac{G_L}{G_H}\Phi\right]-f_H G_H\frac{1}{u_H}\left[1-\frac{f_L}{f_H}\frac{G_L}{G_H}\Phi\right]-f_H G_H\eta(u_H;\sigma)\frac{1}{u_H}\right. \\
& \left.-f_L G_L\Phi\Big[\eta(u_H;\sigma)\frac{1}{u_H}-\eta(u_L;\sigma)\frac{1}{u_L}\Big]\right)\left(-\frac{u_L}{\sigma}f_L G_L\frac{\partial\eta(u_L;\sigma)}{\partial u}\frac{1}{\eta(u_L;\sigma)}[1+\Phi]\right) \\
=& \frac{u_H}{\sigma}f_H G_H\frac{\partial\eta(u_H;\sigma)}{\partial u}\frac{1}{\eta(u_H;\sigma)}\Big[1-\frac{f_L}{f_H}\frac{G_L}{G_H}\Phi\Big] \\
& \times\left(f_L G_L\frac{\partial\eta(u_L;\sigma)}{\partial u}\frac{1}{u_L}[1+\Phi]-f_L G_L\frac{1}{u_L}-f_L G_L\eta(u_L;\sigma)\frac{1}{u_L}[1+\Phi]\right) \\
& -\frac{u_L}{\sigma}f_L G_L\frac{\partial\eta(u_L;\sigma)}{\partial u}\frac{1}{\eta(u_L;\sigma)}[1+\Phi] \\
& \times\left(f_H G_H\frac{\partial\eta(u_H;\sigma)}{\partial u_H}\frac{1}{u_H}\left[1-\frac{f_L}{f_H}\frac{G_L}{G_H}\Phi\right]-f_H G_H\frac{1}{u_H}\left[1-\frac{f_L}{f_H}\frac{G_L}{G_H}\Phi\right]-f_H G_H\eta(u_H;\sigma)\frac{1}{u_H}\right) \\
& -f_L G_L\frac{1}{\sigma}\Phi\Big[\eta(u_H;\sigma)-\eta(u_L;\sigma)\Big]\left(f_L G_L\frac{\partial\eta(u_L;\sigma)}{\partial u}\frac{1}{u_L}[1+\Phi]-f_L G_L\frac{1}{u_L}-f_L G_L\eta(u_L;\sigma)\frac{1}{u_L}[1+\Phi]\right) \\
& +f_L G_L\Phi\Big[\eta(u_H;\sigma)\frac{1}{u_H}-\eta(u_L;\sigma)\frac{1}{u_L}\Big]\frac{u_L}{\sigma}f_L G_L\frac{\partial\eta(u_L;\sigma)}{\partial u}\frac{1}{\eta(u_L;\sigma)}[1+\Phi]
\end{aligned}
$$

$$
\begin{aligned}
&-\frac{\partial^2\Pi}{\partial\sigma\partial u_H}\frac{\partial^2\Pi}{\partial(u_L)^2}+\frac{\partial^2\Pi}{\partial u_H\partial u_L}\frac{\partial^2\Pi}{\partial\sigma\partial u_L}\\
=&-\left\{-\frac{u_H}{\sigma}f_HG_H\frac{\partial\eta(u_H;\sigma)}{\partial u}\frac{1}{\eta_H}+\frac{u_H}{\sigma}f_LG_L\frac{\partial\eta(u_H;\sigma)}{\partial u}\frac{1}{\eta(u_H;\sigma)}\Phi+f_LG_L\frac{1}{\sigma}\Phi\eta(u_H;\sigma)-f_LG_L\frac{1}{\sigma}\Phi\eta(u_L;\sigma)\right\}\\
&\left\{f_LG_L\frac{\partial\eta(u_L;\sigma)}{\partial u}\frac{1}{u_L}[1+\Phi]-f_LG_L\frac{1}{u_L}[1+\Phi]-f_LG_L\eta(u_L;\sigma)\frac{1}{u_L}[1+\Phi]+f_LG_L\frac{\Phi'}{\Delta'}\right\}\\
&+\left\{f_LG_L\frac{1}{u_L}\eta(u_L;\sigma)\Phi-f_LG_L\frac{\Phi'}{\Delta'}\right\}\left\{-\frac{u_L}{\sigma}f_LG_L\frac{\partial\eta(u_L;\sigma)}{\partial u}\frac{1}{\eta(u_L;\sigma)}[1+\Phi]\right\}\\
=&f_L^2G_L^2\left\{\frac{u_H}{\sigma}\frac{f_HG_H}{f_LG_L}\frac{\partial\eta(u_H;\sigma)}{\partial u}\frac{1}{\eta(u_H;\sigma)}\left[1-\frac{f_LG_L}{f_HG_H}\Phi\right]-\frac{1}{\sigma}\Phi\Big[\eta(u_H;\sigma)-\eta(u_L;\sigma)\Big]\right\}\\
&\left\{\frac{\partial\eta(u_L;\sigma)}{\partial u}\frac{1}{u_L}[1+\Phi]-\frac{1}{u_L}[1+\Phi]-\eta(u_L;\sigma)\frac{1}{u_L}[1+\Phi]+\frac{\Phi'}{\Delta'}\right\}\\
&-\frac{u_L}{\sigma}f_L^2G_L^2\left\{\frac{1}{u_L}\eta(u_L;\sigma)\Phi-\frac{\Phi'}{\Delta'}\right\}\frac{\partial\eta(u_L;\sigma)}{\partial u}\frac{1}{\eta(u_L;\sigma)}[1+\Phi]\\
=&f_HG_Hf_LG_L\frac{\partial\eta(u_L;\sigma)}{\partial u}\frac{1}{u_L}[1+\Phi]\frac{u_H}{\sigma}\frac{\partial\eta(u_H;\sigma)}{\partial u}\frac{1}{\eta(u_H;\sigma)}\left[1-\frac{f_LG_L}{f_HG_H}\Phi\right]\\
&-f_HG_Hf_LG_L\eta(u_L;\sigma)\frac{1}{u_L}[1+\Phi]\frac{u_H}{\sigma}\frac{\partial\eta(u_H;\sigma)}{\partial u}\frac{1}{\eta(u_H;\sigma)}\left[1-\frac{f_LG_L}{f_HG_H}\Phi\right]\\
&-f_L^2G_L^2\frac{1}{\sigma}\Phi\frac{1}{u_L}[1+\Phi]\Big[\eta(u_H;\sigma)-\eta(u_L;\sigma)\Big]\left\{\frac{\partial\eta(u_L;\sigma)}{\partial u}-1-\eta(u_L;\sigma)\right\}\\
&-f_HG_Hf_LG_L\frac{1}{u_L}[1+\Phi]\frac{u_H}{\sigma}\frac{\partial\eta(u_H;\sigma)}{\partial u}\frac{1}{\eta(u_H;\sigma)}\left[1-\frac{f_LG_L}{f_HG_H}\Phi\right]\\
&-f_L^2G_L^2[1+\Phi]\frac{1}{\sigma}\frac{\partial\eta(u_L;\sigma)}{\partial u}\Phi\\
&+f_LG_L\frac{\Phi'}{\Delta'}\left\{\frac{u_L}{\sigma}f_LG_L\frac{\partial\eta(u_L;\sigma)}{\partial u}\frac{1}{\eta(u_L;\sigma)}[1+\Phi]+\frac{u_H}{\sigma}f_HG_H\frac{\partial\eta(u_H;\sigma)}{\partial u}\frac{1}{\eta(u_H;\sigma)}\left[1-\frac{f_LG_L}{f_HG_H}\Phi\right]\right\}\\
&-f_L^2G_L^2\frac{\Phi'}{\Delta'}\frac{1}{\sigma}\Phi\Big[\eta(u_H;\sigma)-\eta(u_L;\sigma)\Big]
\end{aligned}
$$

$$
\begin{aligned}
&-\frac{\partial^2\Pi}{\partial(u_H)^2}\frac{\partial^2\Pi}{\partial\sigma\partial u_L}+\frac{\partial^2\Pi}{\partial\sigma\partial u_H}\frac{\partial^2\Pi}{\partial u_H\partial u_L}\\
&=-f_LG_L\frac{\partial\eta(u_L;\sigma)}{\partial\sigma}\frac{1}{\eta(u_L;\sigma)}[1+\Phi]\Bigg\{f_HG_H\frac{\partial\eta(u_H;\sigma)}{\partial u_H}\frac{1}{u_H}\left[1-\frac{f_LG_L}{f_HG_H}\Phi\right]-f_HG_H\frac{1}{u_H}\left[1-\frac{f_LG_L}{f_HG_H}\Phi\right]\\
&-f_HG_H\eta(u_H;\sigma)\frac{1}{u_H}-f_HG_H\eta(u_H;\sigma)\frac{1}{u_H}\frac{f_L}{f_H}\frac{G_L}{G_H}\Phi+f_LG_L\frac{\Phi'}{\Delta'}\Bigg\}\\
&+f_LG_L\frac{1}{u_L}\eta(u_L;\sigma)\Phi\left\{f_HG_H\frac{\partial\eta(u_H;\sigma)}{\partial\sigma}\frac{1}{\eta(u_H;\sigma)}\left[1-\frac{f_L}{f_H}\frac{G_L}{G_H}\Phi\right]+f_LG_L\frac{1}{\sigma}\Phi\Big[\eta(u_H;\sigma)-\eta(u_L;\sigma)\Big]\right\}\\
&-f_LG_L\frac{\Phi'}{\Delta'}\left\{f_HG_H\frac{\partial\eta(u_H;\sigma)}{\partial\sigma}\frac{1}{\eta(u_H;\sigma)}\left[1-\frac{f_L}{f_H}\frac{G_L}{G_H}\Phi\right]+f_LG_L\frac{1}{\sigma}\Phi\Big[\eta(u_H;\sigma)-\eta(u_L;\sigma)\Big]\right\}\\
&=-f_LG_L\frac{\partial\eta(u_L;\sigma)}{\partial\sigma}\frac{1}{\eta(u_L;\sigma)}[1+\Phi]\Bigg\{f_HG_H\frac{\partial\eta(u_H;\sigma)}{\partial u_H}\frac{1}{u_H}\left[1-\frac{f_LG_L}{f_HG_H}\Phi\right]\\
&-f_HG_H\frac{1}{u_H}\left[1-\frac{f_LG_L}{f_HG_H}\Phi\right]\\
&-f_HG_H\eta(u_H;\sigma)\frac{1}{u_H}-f_HG_H\eta(u_H;\sigma)\frac{1}{u_H}\frac{f_L}{f_H}\frac{G_L}{G_H}\Phi\Bigg\}\\
&+f_LG_L\frac{1}{u_L}\eta(u_L;\sigma)\Phi\left\{f_HG_H\frac{\partial\eta(u_H;\sigma)}{\partial\sigma}\frac{1}{\eta(u_H;\sigma)}\left[1-\frac{f_L}{f_H}\frac{G_L}{G_H}\Phi\right]+f_LG_L\frac{1}{\sigma}\Phi\Big[\eta(u_H;\sigma)-\eta(u_L;\sigma)\Big]\right\}\\
&-f_LG_L\frac{\Phi'}{\Delta'}\Bigg\{f_HG_H\frac{\partial\eta(u_H;\sigma)}{\partial\sigma}\frac{1}{\eta(u_H;\sigma)}\left[1-\frac{f_L}{f_H}\frac{G_L}{G_H}\Phi\right]+f_LG_L\frac{\partial\eta(u_L;\sigma)}{\partial\sigma}\frac{1}{\eta(u_L;\sigma)}[1+\Phi]\\
&+f_LG_L\frac{1}{\sigma}\Phi\Big[\eta(u_H;\sigma)-\eta(u_L;\sigma)\Big]\Bigg\}
\end{aligned}
$$

Using the IC_H relation $u_H-u_L=\Delta(q_L)$, we have

$$
\Delta'(q_L)\frac{dq_{L,2}}{d\sigma}=\frac{du_{H,2}}{d\sigma}-\frac{du_{L,2}}{d\sigma}=-\frac{u_{L,2}}{\sigma}\frac{\eta_L\Phi\Big[1+\eta_L+\Phi\Big]}{-(\eta_L)^2(\Phi)^2-\mathcal{E}^{\Phi}_{u_L}\eta_L(\Phi)^2-\mathcal{E}^{\eta}_{u,L}\mathcal{E}^{\Phi}_{u_L}\Phi-\mathcal{E}^{\eta}_{u,L}\mathcal{E}^{\Phi}_{u_L}(\Phi)^2+\mathcal{E}^{\Phi}_{u_L}\Phi\Big[1+\eta_L+\Phi\Big]}
$$

$$\frac{\sigma}{u_{H,2}}\frac{du_{H,2}}{d\sigma} = \frac{-\eta_L\Phi\left[1+\eta_L+\Phi\right]-(\eta_L)^2(\Phi)^2-\mathcal{E}^{\Phi}_{u_L}\eta_L(\Phi)^2-\mathcal{E}^{\eta}_{u,L}\mathcal{E}^{\Phi}_{u_L}\Phi-\mathcal{E}^{\eta}_{u,L}\mathcal{E}^{\Phi}_{u_L}(\Phi)^2}{-(\eta_L)^2(\Phi)^2-\mathcal{E}^{\Phi}_{u_L}\eta_L(\Phi)^2-\mathcal{E}^{\eta}_{u,L}\mathcal{E}^{\Phi}_{u_L}\Phi-\mathcal{E}^{\eta}_{u,L}\mathcal{E}^{\Phi}_{u_L}(\Phi)^2+\mathcal{E}^{\Phi}_{u_L}\Phi\left[1+\eta_L+\Phi\right]}$$

$$=1-\frac{\left[\eta_L+\mathcal{E}^{\Phi}_{u_L}\right]\Phi\left[1+\eta_L+\Phi\right]}{-(\eta_L)^2(\Phi)^2-\mathcal{E}^{\Phi}_{u_L}\eta_L(\Phi)^2-\mathcal{E}^{\eta}_{u,L}\mathcal{E}^{\Phi}_{u_L}\Phi-\mathcal{E}^{\eta}_{u,L}\mathcal{E}^{\Phi}_{u_L}(\Phi)^2+\mathcal{E}^{\Phi}_{u_L}\Phi\left[1+\eta_L+\Phi\right]}$$

$$\frac{\sigma}{u_{L,2}}\frac{du_{L,2}}{d\sigma} = \frac{-(\eta_L)^2(\Phi)^2-\mathcal{E}^{\Phi}_{u_L}\eta_L(\Phi)^2-\mathcal{E}^{\eta}_{u,L}\mathcal{E}^{\Phi}_{u_L}\Phi-\mathcal{E}^{\eta}_{u,L}\mathcal{E}^{\Phi}_{u_L}(\Phi)^2}{-(\eta_L)^2(\Phi)^2-\mathcal{E}^{\Phi}_{u_L}\eta_L(\Phi)^2-\mathcal{E}^{\eta}_{u,L}\mathcal{E}^{\Phi}_{u_L}\Phi-\mathcal{E}^{\eta}_{u,L}\mathcal{E}^{\Phi}_{u_L}(\Phi)^2+\mathcal{E}^{\Phi}_{u_L}\Phi\left[1+\eta_L+\Phi\right]}$$

$$=1-\frac{\mathcal{E}^{\Phi}_{u_L}\Phi\left[1+\eta_L+\Phi\right]}{-(\eta_L)^2(\Phi)^2-\mathcal{E}^{\Phi}_{u_L}\eta_L(\Phi)^2-\mathcal{E}^{\eta}_{u,L}\mathcal{E}^{\Phi}_{u_L}\Phi-\mathcal{E}^{\eta}_{u,L}\mathcal{E}^{\Phi}_{u_L}(\Phi)^2+\mathcal{E}^{\Phi}_{u_L}\Phi\left[1+\eta_L+\Phi\right]}$$

So, $\frac{dq_{L,2}}{d\sigma} < 0$. Moreover, since $\frac{\sigma}{u_{H,2}}\frac{du_{H,2}}{d\sigma} < \frac{\sigma}{u_{L,2}}\frac{du_{L,2}}{d\sigma} < 1$we have $\frac{dG(u_L;\sigma)}{d\sigma} < 0$.

Chapter 3

Imperfect competition in insurance markets

3.1 Introduction

Adverse selection and imperfect competition are two key problems affecting many private insurance markets and are pertinent to consumers, insurers and policy-makers alike. Since the seminal work of Rothschild and Stiglitz (1976) (RS), asymmetric information in individuals' risk has been considered a key ingredient in modelling insurance markets and its presence is typically predicted to lead to inefficient insurance provision in general and adverse selection in particular. These theoretical predictions and their welfare implications have been studied extensively in the subsequent empirical literature. In contrast, the issue of imperfect competition has only recently gained attention in the context of insurance markets. Building on initial evidence highlighting high concentration and market power amongst private insurers (for example Dafny (2010); Dafny et al. (2012)), a growing body of work in the empirical industrial organisation (IO) literature has documented price-discriminatory behaviour in these imperfectly competitive insurance markets and studied their welfare consequences.

This paper proposes a contract-theoretic model of an imperfectly competitive insurance market and uses it to study how the degree of competition between insurers affects equilibrium allocations. I consider a setting where individuals are heterogeneous in two dimensions: their risk of incurring a loss and a "horizontal" preference for particular insurers. Imperfect competition thus enters the model through horizontally-differentiated insurers. We assume both dimensions of heterogeneity are private information and insurers strategically offer menus of contracts which specify coverage-premium pairs. In relation to the existing literature, the present analysis embeds a RS-type insurance model in a in a Hotelling environmentwith horizontal differentiation between firms.

We show that the qualitative features of the RS/Stiglitz equilibrium allocations obtain with levels of competition between the perfectly competitive and monopoly limits. Insurers always seek to separate the two risk-groups by offering menus consisting of a full-insurance contract at a higher premium-rate targeted at high-risk individuals and a partial-insurance contract at a lower premium-rate intended for low-risks. In contrast to the perfectly competitive case, equilbrium coverage obtained by low-risk individuals and consumer surplus for both risk groups are characterised by conditions which capture intensive and extensive margin (from switching insurers) tradeoffs in profit.

Our setup also allows us to study how equilibrium allocations vary when the degree of competition changes. In this regard, we find that consumer surplus is strictly increasing in the degree of competition. More interestingly, coverage provided to low-risks decreasing in the degree of competition. Intuitively, insurers can use either differential pricing or insurance rationing to screen risks. Rationing reduces the surplus from trade thereby reducing the total potential profit that can be made from an insurance contract. An insurer would therefore prefer to substitute away from rationing. However, strong competition forces insurers to price close to expected costs and limits the use of differential pricing for screening. Decreasing the intensity of competition between insurers thus provides insurers with more flexibility in setting prices above cost and this, in turn, allows them to substitute away from rationing coverage towards differential prices.

Finally, our analysis also shows that for each level of market competition, equilibrium exists whenever there is a sufficiently large proportion of high-risk individuals. Symmetrically, for a given risk distribution, there is also a threshold level of market power above which a pure strategy equilibrium is exists. Together, these results imply that relaxing the assumption of perfect competition to allow for imperfectly competitive insurer can help to address the issue of equilibrium non-existence in the standard Rothschild-Stiglitz setting.

3.1.1 Related literature

This paper builds on the theoretical literature studying adverse selection in insurance markets stemming from the seminal work of Rothschild and Stiglitz (1976); Wilson (1977) for the perfectly competitive case and Stiglitz (1977); Chade and Schlee (2012) for an insurance monopoly. the use of the Hotelling environment to model an imperfectly competitive insurance market builds on previous work by Jack (2006); Olivella and Vera-Hernández (2007) who study horizontally-differentiated health plans and more recently by Bijlsma et al. (2014).

Nonlinear pricing with imperfect competition
The literature on the nonlinear pricing problem with imperfect competition starts from **?** and **?**. Similar to the present analysis, these papers embed a firm's problem of contracting with unobservable types (a continuum in this case) within a Hotelling environment and consider how equilibrium allocations within the market are affected by whether or not there is an active participation margin. This issue of market coverage is studied in more detail with the same continuum-type model in Yang and Ye (2008) and with two discrete types in Shen et al. (2016).

However, a key difference between the nonlinear pricing problem and the insurance screening problem which we consider is that both willingness to pay and cost are type-dependent. This double type-dependence is a feature of the models of Villas-Boas and Schmidt-Mohr (1999) and Bénabou and Tirole (2016) whose analytical approach ours most closely resembles. The former embeds a credit market problem in the spirit of Bester et al. (1985) within the Hotelling framework and studies equilibrium credit rationing. Bénabou and Tirole (2016) studies the equilibrium of a labour market screening problem embedded in a Hotelling environment and assumes co-located outside options which preclude the case with an active participation margin. The present analysis adopts the approach of Bénabou and Tirole (2016), by similarly assuming co-located outside options. This allows us to isolate the effect of competition on intensive margin allocations.

3.2 Model

3.2.1 Timing and risk

There is a unit mass of individuals, each of whom starts with wealth, w, and faces a risk of incurring an exogenous monetary loss, $l > 0$. An insurance contract indemnifies against a fraction, $q \geq 0$, of the loss and costs premium $p \geq 0$. There is a single insurer a menu of coverage-premium contracts. Individuals can either purchase insurance from the two insurers or remain uninsured. We focus on the setting with exclusive contracting and full commitment with the following timing of events: the insurer offers a menu of contracts; individuals observe these offers and choose either to purchase insurance or remain uninsured; losses are realised and payouts are made.

3.2.2 Individuals' preferences

Individuals are heterogeneous and differ along two dimensions denoted by $(\theta, x) \in \Theta \times [0,1]$. We consider the case with $\Theta = \{\theta_L, \theta_H\}$ where $\theta_L < \theta_H$. f_i is the proportion of θ_i, $i = H, L$, in the population, with $f_H + f_L = 1$. For the sake of exposition, we refer to θ as an individual's risk and x their "brand preferences".

Risk preferences Let $U(q,p;\theta)$ denote the expected utility of an individual with risk θ under contract (q,p). Let $v(q;\theta)$ be defined implicitly by $U(q, v(q;\theta);\theta) = U(0,0;\theta)$. Conditional on purchasing insurance, $v(q;\theta)$ is the maximum amount that a θ-risk individual is willing to pay for q-level of coverage and is a money-metric representation of an individual's risk preferences. Note that by definition, $v(0;\theta) = 0$. We call $v(q;\theta)$ the "willingness to pay" for q by θ-risks and hereafter work with $v(q;\theta)$ instead of U in our analysis. Assume risk preferences are such that $v(q;\theta)$ has the following properties.

Assumption 5. *Assume $v(q;\theta)$ is twice continuously differentiable and satisfies*

1. $v_q(q;\theta) \geq 0$, $v_{qq}(q;\theta) < 0$ *(concavity)*
2. $v_\theta(q;\theta) \geq 0$ *(risk-dependence)*
3. $v_{q\theta}(q;\theta) > 0$ *(single-crossing in (q,θ))*

Assumption 5 requires individuals' willingness to pay to be increasing and strictly concave in q (Point 1). Furthermore, it assumes that willingness to pay for each *level* of coverage is non-decreasing in individuals' risk (Point 2) and that the *marginal* willingness to pay for each additional unit of coverage is increasing in risk (Point 3). Below, we show that Assumption 5 is satisfied in the most commonly analysed insurance environments.

Brand preferences Let k index insurers and let $u_i^k \equiv v(q_i^k;\theta_i) - p_i^k$ denote the gross surplus enjoyed by a θ_i-risk individual from (q_i^k, p_i^k). We model horizontal differentiation using the Hotelling-Salop framework. Specifically, let (q_i^0, p_i^0) and (q_i^1, p_i^1) be the contracts offered by Insurer-0 and Insurer-1 respectively. Conditional becoming insured, a θ_i-individual prefers purchasing (q_i^0, p_i^0) from Insurer-0 over (q_i^1, p_i^1) from Insurer-1 if and only if

$$v(q_i^0;\theta_i) - p_i^0 - \tau x \geq v(q_i^1;\theta_i) - p_i^1 - \tau(1-x) \iff u_i^0 - \tau x \geq u_i^1 - \tau(1-x) \qquad (3.1)$$

We assume that x is independent of θ and is uniformly distributed on $[0,1]$. x captures an individual's preference over insuring with Insurer-0 versus Insurer-1 and $\tau \geq 0$ captures the intensity of these preferences. Note also that since Condition **??** is in wealth space, τx is money-metric like p and $v(q;\theta)$.

Autarky payoff We model autarky payoffs following Bénabou and Tirole (2016). Specifically, assume that an individual with (θ_i, x), purchases insurance from Insurer-0 if and only if

$$u_i^0 - \tau x \geq \max\Big\{u_i^1 - \tau(1-x),\ -\tau x,\ -\tau(1-x)\Big\} \qquad (3.2)$$

The last two terms in the max-operator on the right-hand side implies that individuals have autarky outside options with the same *non-contract* characteristics as each of the two insurers. Intuitively, this means an individual takes up insurance if and only if the gains from their preferred option within the insurance market is greater than their preferred option in autarky.[1] As discussed in Bénabou and Tirole (2016) this approach ensures that brand preferences do not impact the decision to participate directly.

[1] For example, consider two financial institutions. Each financial institution can provide both formal insurance and also saving deposits. The latter serves as our autarky outside option. In this example, x represents an individual's preference for a particular financial institution, while y could represent the hassle cost associated with purchasing a new insurance contract. Consider an individual with $x < 0.5$ (This could arise if, for example, they already have an existing banking relationship with Firm-0). Then, conditional on choosing to purchase insurance, the individual would choose to contract with Firm-0, while conditional on remaining uninsured, the individual simply continues saving (i.e. self insuring) with the same firm.

3.2.3 Insurer's profit

Insurers' profit depends only on the contract offered and the insured's risk. Let $c(q;\theta)$ be the cost of providing coverage q to an individual of risk θ.

Assumption 6. *Assume $c(q;\theta)$ is twice continuously differentiable and satisfies*

1. $c(0;\theta) = 0$
2. $c_q(q;\theta) > 0$, $c_{qq}(q;\theta) \geq 0$ *(convexity)*
3. $c_\theta(q;\theta) > 0$ *for* $q > 0$ *(risk-dependence)*
4. $c_{q\theta}(q;\theta) > 0$

Point 1 assumes zero costs when no coverage is provided while Point 2 assumes that costs are strictly increasing and convex in coverage. The third and fourth points imply that the total (Point 3) and marginal cost (Point 4) of providing insurance is increasing in risk. Points 3 and 4 distinguish our setting from the standard nonlinear product pricing problem (in the vein of Mussa-Rosen) where costs are independent of consumers' "type".

Surplus from insurance We denote the surplus from providing insurance coverage, q, to an individual with θ by $s(q;\theta) \equiv v(q;\theta) - c(q;\theta)$. The insurer's profit from providing (q_i, p_i) to a θ_i-individual is $s(q_i;\theta_i) - u_i$. Our definition of $s(q;\theta)$ does not account for brand preferences τx, as is standard in this literature.

By assumptions in Assumptions 5 and 6, we know that $s(q;\theta)$ is strictly concave and hence has a unique maximiser $q_i^* = \arg\max_q s(q;\theta_i)$. For brevity, we let s_i^* denote the maximised surplus for each i. In addition to Assumptions 5 and 6, we further assume that $s(q;\theta)$ satisfies the following property.

Assumption 7. $q_i^* = q^* > 0$ *for each* $\theta_i \in \Theta$.

Assumption 7 says that the surplus-maximising level of coverage exists, is positive and does not depend on individuals' risk. It is satisfied in standard models of insurance contracting with unobservable risk. However, this assumption may be violated in models where risk aversion is correlated with risk and in other behavioural models with non-standard preferences.

Incentive compatibility

Restricting the analysis to deterministic contracts, the revelation principle applies and allows us to consider direct revelation mechanisms without loss of generality.[2] Unless otherwise noted, we hereafter focus on Insurer-0 and drop superscripts.

Define $\Delta(q) \equiv v(q;\theta_H) - v(q;\theta_L)$. Under Assumption 5, $\Delta(q)$ is positive, strictly increasing and convex for all $q \geq 0$ and equal to zero at $q = 0$. Using our definition of u_i and $\Delta(q)$, the incentive compatibility constraints can be rewritten as

$$\begin{aligned} IC_H: &\quad u_H \geq u_L + \Delta(q_L) \\ IC_L: &\quad u_L \geq u_H - \Delta(q_H) \end{aligned}$$

$u_L + \Delta(q_L)$ is the surplus that a high-risk individual obtains if they were to purchase the contract, (q_L, p_L). $\Delta(q_L)$ is thus the "information rent" that is required by high-risk individuals to ensure they purchase (q_H, p_H) instead. A symmetric interpretation applies for IC_L.

Market shares

For a given u^1 offered by Insurer-1, the share of individuals contracting with Insurer-0 when the latter offers u is $H(u;u^1,\tau) = \min\{1, \max\{1/2 + (u - u^1)/2\tau, 0\}\}$.

Equilibrium definition

Given Insurer-1's menu of contracts, $C^1 \equiv \{(q_i^1; u_i^1)\}_{i=H,L}$, Insurer-0 chooses $C^0 \equiv \{(q_i^0; u_i^0)\}_{i=H,L}$ to maximise

$$\Pi(C^j;C^{-j}) = \sum_{i=H,L} H(u_i^0; u_i^1, \tau)[s(q_i^0;\theta_i) - u_i^0] \quad s.t. \quad (IC_H),\, (IC_L)\; and\; q_i^0 \geq 0, u_i^0 \geq 0 \quad \forall i \in H, L$$

We consider symmetric pure strategy Nash equilibria. An equilibrium in pure strategies in our market is a menu, C^e, such that C^e solves Insurer-0's problem given $C^1 = C^e$, and C^e solves Insurer-1's problem given $C^0 = C^e$.

[2]See **?** for a discussion for how this might not be the case with stochastic contracts and the argument in defence of restricting to deterministic contracts.

3.3 Benchmark allocations

We first characterise the perfectly competitive benchmark which abstracts from brand and participation preferences.

3.3.1 Perfectly competitive allocation

Let $C^{LCS} \equiv \{(u_i^0, q_i^0)\}_{i=H,L}$ denote the least-cost separating (LCS) allocation where

$$u_H^0 = s_H^*, \quad q_H^0 = q^*, \quad u_L^0 = s(q_L^0; \theta_L) \tag{3.3}$$

$$s_H^* - s(q_L^0; \theta_L) - \Delta(q_L^0) = 0 \tag{3.4}$$

The first line says that high-risk individuals receive the surplus-maximising level of coverage and contracts for high and low-risks make zero-profit individually. The second line says that coverage for low-risks is chosen to satisfy the IC_H constraint and is in general below the surplus-maximising level.

Lemma 15. *Suppose Assumptions 5, 6, 7 hold. There exists a threshold, $\phi^0 > 0$ such that if $\frac{f_H}{f_L} \geq \phi^0$, a unique perfectly competitive equilibrium exists. Moreover, the equilibrium allocation is given by $C^{LCS} \equiv \{(u_i^0, q_i^0)\}_{i=H,L}$.*

If there is a sufficiently large proportion of high-risk individuals, a perfectly competitive equilibrium in pure strategies exists. Contracts for each risk-type earn zero profit and high-risk individuals are offered the surplus-maximising level of coverage, q^*. Low-risk individuals' coverage is rationed and set to ensure separation of risks by satisfying the IC_H constraint. The threshold which ensures existence is given by $\phi^0 = \Phi(q_L^0)$ where q_L^0 is that LCS level of coverage for low-risk individuals. Recall that $\Phi(q)$ represents the ratio of marginal surplus gain over the marginal informational rent loss from raising the level of low-risk coverage, q. The existence condition, $\frac{f_H}{f_L} \geq \phi^0$, thus requires that starting from q_L^0, the loss from increasing low-risks' coverage exceeds its associated gains. ϕ^0 is below the ϕ^∞ since the former evaluates the marginal gain-loss ratio at a positive level of coverage and the latter at zero coverage.

3.3.2 Monopoly allocation

Lemma 16. *Suppose Assumptions 5, 6, 7 hold. A solution to the monopoly insurance problem always exists. Moreover there exists a threshold, $\phi^{\infty} > 0$ such that*

- *For $\frac{f_H}{f_L} \leq \phi^{\infty}$, the solution to the problem is $C^{monop} = \{(\Delta(q_L^{\infty}), q^*), (0, q_L^{\infty})\}$ where $\Phi(q_L^{\infty}) = \frac{f_H}{f_L}$*
- *For $\frac{f_H}{f_L} > \phi^{\infty}$, the solution is $C^{monop} = \{(0, q^*), (0, 0)\}$.*

This monopoly benchmark is discussed in detail in the companion chapter on monopoly insurance and take-up.

3.4 Insurance market with horizontal differentiation

3.4.1 Existence of equilibrium in pure strategies

Proposition 4. *Suppose Assumptions 5, 6 and 7 hold.*

- *For each $\tau \geq 0$, there exists a threshold proportion of high-to-low risks, $\underline{\phi} > 0$, such that an equilibrium in pure strategies exists if $\frac{f_H}{f_L} \geq \underline{\phi}$. When $\tau \to 0$, this threshold approaches ϕ^0 as defined in the perfectly competitive benchmark.*
- *For each $\frac{f_H}{f_L} \geq 0$, there exists a threshold level of competition, $\underline{\tau} \geq 0$, such that an equilibrium in pure strategies exists if $\tau \geq \underline{\tau}$. When $\frac{f_H}{f_L} \geq \underline{\phi}$, the threshold is equal to zero.*

Part 1 of Proposition 4 says that for each level of market competition, a symmetric equilibrium exists whenever there is a sufficiently large proportion of high-risk individuals. The result generalises the existence condition of the perfectly competitive insurance setting to the case with imperfectly competitive duopoly setting. The result is based on the same logic of profitable cross-subsidising deviations as the existence proof for the perfectly competitive case. The main difference for the case with $\tau > 0$, is that the focal deviation now corners the high-risk segment of the market and uses increased profit from low-risk contracts in order to compensate for the cost of capturing all high-risk individuals. Such a deviation generates more profit than the symmetric interior menu if there are sufficiently few high-risk individuals. This ensures that the surplus gains from improving low-risks' coverage can compensate for the profit loss from high-risks' contracts. A corollary of this is that if the proportion of high-risks is high enough to ensure the existence of a pure strategy equilibrium in the limit as the market tends to perfect competition, then an equilibrium exists for any strictly positive degree of horizontal differentiation.

Part 2 of Proposition 4 is symmetric to the first part. It says that taking the risk distribution in the population as given, we can find a threshold level of market power above which a pure strategy equilibrium is exists. Intuitively, for a given $\frac{f_H}{f_L}$, higher τ increases the size of profit lost in order to capture the high-risk segment. This, in turn, means there is less scope for the deviation to be profitable. A corollary of this is that for any given distribution of risk, a pure strategy equilibrium can exist provided competition between insurers is sufficiently weak.

3.4.2 Characterisation

Proposition 5. *Suppose Assumptions 5, 6 and 7 hold. Assume* $\frac{f_H}{f_L} \geq \phi^0$*. Then a symmetric equilibrium in pure strategies exists for all* $\tau \geq 0$*. Let the equilibrium menu be denote* $C^e = \{(u_H, q_H), (u_L, q_L)\}$*.* C^e *is such that* $u_H = u_L + \Delta(q_L)$*,* $q_H = q^*$ *and*

$$\frac{1}{\tau}[s_H^* - u_H] - 1 + \frac{f_L}{f_H}\Phi(q_L) = 0 \tag{3.5}$$

$$f_H[s_H^* - u_H] + f_L[s(q_L; \theta_L) - u_L] = \tau \tag{3.6}$$

Proposition 5 says that an equilibrium with imperfect competition always provides high-risk individuals with "full coverage" and is such that high-risk individuals are left indifferent between their own contract and that intended for low-risks, i.e. the incentive constrant IC_H is binding. These familiar properties arise from standard arguments which apply to almost all contracting problems with unobservable "types". On the other hand, with imperfect competition u_H, u_L and q_L are jointly chosen to trade off extensive margin profit changes for each risk group against intensive margin changes in total profit.

Condition 3.5 highlights the intensive and extensive margin tradeoffs when u_H is adjusted. The first term is the per-contract extensive margin gain in profit from increasing u_H. The second term is the per-contract intensive-margin loss in profit from increasing u_H. The third term captures the intensive-margin gain in low-risk profit since increasing u_H relaxes the IC_H constraint and allows one to increase q_L. Similarly, Condition (3.6) says that for a given u_H, u_L equates the extensive margin decrease in profit from low-risk individuals (the second term divided by τ) with the intensive margin changes in *total* profit. The intensive margin effects consist of a direct effect from the change in share of surplus left to consumers and an indirect effect, through the change in size of surplus from each trade. In equilibrium, this intensive margin effect is equal to the first term, divided by τ. Note that Condition (3.6) implies that an insurer's equilibrium total expected profit is strictly increasing in τ and approaches zero as τ approaches zero.

Corollary 6. *Let $C^e = \{(u_H, q_H), (u_L, q_L)\}$ denote an equilibrium menu. Then u_H and u_L are strictly decreasing in τ while q_L is strictly increasing in τ*

Corollary 6 says that consumer surplus is strictly increasing and the level of coverage provided to low-risks is strictly decreasing in the degree of competition. Intuitively, as the intensity of brand preferences becomes stronger, each insurer is able to capture a larger share of the surplus from trade, $s(q_i;\theta_i)$, without losing market share. As long as consumer surplus is above zero, the size of the entire market is not affected by the shift in surplus allocation either. Therefore, we find that increasing τ, which can be interpreted as weakening competition, leads to decreasing consumer surplus. In contrast, coverage provided to low-risks is increasing in τ. Intuitively, insurers have two instruments to achieve equilibrium risk separation: differential pricing (i.e. adjusting u_i) and insurance rationing (i.e. adjusting q_L). Rationing is socially inefficient as it reduces the surplus from trade, $s(q_L;\theta_L)$. This distortion is also undesirable for insurers as it also reduces the total potential surplus that can be captured. Strong competition forces insurers to price close to expected costs and limits the use of differential pricing for screening. Decreasing the intensity of competition between insurers thus provides them with more flexibility in setting prices above cost and this, in turn, allows them to "substitute" away from rationing coverage towards differential prices.

Finally, taking the limit as τ goes to zero, the allocation in our model approaches the perfectly competitive benchmark allocation C^{comp}. Conversely, our model's allocation approaches the full take-up monpoly benchmark, C^{monop}, in the limit as τ becomes arbitrarily large

Corollary 7. *Let $C \equiv \{(u_H, q_H), (u_L, q_L)\}$ denote the equilibrium menu. Then $C = C^{comp}$ in the limit as $\tau \to 0$ and $C = C^{monop}$ in the limit as $\tau \to +\infty$.*

3.5 Conclusion

In this paper, I proposed and analysed a theoretical model of an imperfectly competitive insurance market with horizontally-differentiated insurers. As with canonical Rothschild-Stiglitz family of models, an equilibrium in our setting has the highest-risk individuals obtaining full insurance coverage and low-risks partial insurance coverage. Equilbrium coverage obtained by low-risk individuals and consumer surplus for both risk groups are characterised by conditions which capture intensive and extensive margin (from switching insurers) tradeoffs in profit. Analysis of this model shows that consumer surplus is, unsurprisingly, decreasing as the degree of competition decreases. In contrast, the degree of "rationing" of low-risk coverage is decreasing as the degree of competition decreases. Put differently, conditional on full take-up of insurance by all individuals, an increase in insurers' market power can improve the efficiency of insurance allocations. Finally, the present analysis also shows that for each level of market competition, equilibrium exists whenever there is a sufficiently large proportion of high-risk individuals. Symmetrically, for a given risk distribution, there is also a threshold level of market power above which a pure strategy equilibrium is exists. Together, these results imply that relaxing the assumption of perfect competition to allow for imperfectly competitive insurer can help to address the issue of equilibrium non-existence in the standard Rothschild-Stiglitz setting.

sectionProof of Lemma 15 Let the "least-cost separating" (LCS) allocation, $(u_H^c, q_H^c), (u_L^c, q_L^c)$, be given by $q_H^c = q^*$ and

$$s_H^* - u_H^c = 0 \tag{3.7}$$
$$s(q_L^c; \theta_L) - u_L^c = 0 \tag{3.8}$$
$$u_H^c - u_L^c - \Delta(q_L^c) = 0 \tag{3.9}$$

The LCS allocation is interim efficient if and only if it solves

$$\begin{aligned}
\max_{u_H, q_H, u_L, q_L} \quad & u_L \\
s.t. \quad & u_H - u_L - \Delta(q_L) \\
& u_L - u_H + \Delta(q_H)] \\
& f_H[s(q_H;\theta_H) - u_H] + f_L[s(q_L;\theta_L) - u_L] \\
& u_H \geq u_H^c = s_H^*
\end{aligned}$$

Let χ be the Lagrange multiplier associated with the break-even constraint and ζ be the multiplier associated with $s_H^* \geq u_H$. The first-order conditions of this problem are

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial u_H} =& \lambda_H - \lambda_L - \chi f_H + \zeta \\
\frac{\partial \mathcal{L}}{\partial u_L} =& 1 - \lambda_H + \lambda_L - \chi f_L \\
\frac{\partial \mathcal{L}}{\partial q_H} =& \lambda_L \Delta'(q_H) + \chi f_H s_q(q_H; \theta_H) \\
\frac{\partial \mathcal{L}}{\partial q_L} =& -\lambda_H \Delta'(q_L) + \chi f_L s_q(q_L; \theta_L)
\end{aligned}$$

First note that the break-even constraint must be binding as otherwise both u_H and u_L can be increased, leaving the IC constraints unaltered and increasing the objective. By the usual arguments, we cannot have $\lambda_H > 0$ and $\lambda_L > 0$ simultaneously. Suppose $\lambda_H = 0$ and $\lambda_L = 0$. Then we have $q_H = q_L = q^*$ and also $\zeta > 0$. Note that both IC constraints hold with equality. Substituting into the break-even constraint, we have $f_H[s_H^* - s_H^*] + f_L[s_L^* - u_L] = f_L[s_L^* - u_H + \Delta(q^*)] = f_L[s_L^* - s_H^* + \Delta(q^*)] = f_L[c(q^*; \theta_H) - c(q^*; \theta_L)] > 0$, a contradiction.

Next, suppose $\lambda_L > 0$ and $\lambda_H = 0$, then we must have $\zeta > 0$ and $q_H > q^* = q_L$. Substituting into the break-even constraint $f_H[s_H^* - s_H^*] + f_L[s_L^* - u_L] = f_L[s_L^* - u_L] \geq f_L[s_L^* - s_H^* + \Delta(q^*)] = f_L[c(q^*; \theta_H) - c(q^*; \theta_L)] > 0$. The second-last inequality obtains from the IC_H condition which implies $u_H \geq u_L + \Delta(q^*) \iff -u_L \geq -s_H^* + \Delta(q^*)$. We thus conclude that $\lambda_H > 0$ and $\lambda_L = 0$ so $q_H = q^* > q_L$.

Case 1: No cross-subsidisation

If $\zeta > 0$, the binding break-even constraint and cross-subsidisation constraints yield $s(q_L; \theta_L) - u_L$. This implies that the interim efficient allocation coincides with the LCS allocation. Furthermore, the F.O.C.s read

$$\lambda_H - \chi f_H + \zeta = 0 \tag{3.10}$$

$$1 - \lambda_H - \chi f_L = 0 \tag{3.11}$$

$$-\lambda_H \Delta'(q_L) + \chi f_L s_q(q_L; \theta_L) = 0 \tag{3.12}$$

Substituting (3.10) and (3.12) yields

$$\zeta - \chi f_L \left[\frac{f_H}{f_L} - \Phi(q_L) \right] = 0 \iff \Phi(q_L) < \frac{f_H}{f_L}$$

Since $\frac{f_H}{f_L} = \Phi(q_L^\infty)$ and since $\Phi(q)$ is strictly decreasing, we have $q_L > q_L^\infty$.

Case 2: Cross-subsidisation

Now, suppose $\zeta = 0$, then the first two F.O.C.s yield $\chi = 1$ and $\lambda_H = f_H$. Using these findings to substitute out χ and λ_H, we have that for $\zeta = 0$, the solution is given by

$$u_H - u_L - \Delta(q_L) = 0 \tag{3.13}$$

$$f_H[s_H^* - u_H] + f_L[s(q_L;\theta_L) - u_L] = 0 \tag{3.14}$$

$$\Phi(q_L) = \frac{f_H}{f_L} \tag{3.15}$$

Substituting and rearranging the first two equations, yields $u_H = f_H s_H^* + f_L s(q_L;\theta_L) + f_L\Delta(q_L)$. Define

$$\Gamma(f_L) \equiv -f_L s_H^* + f_L s(q_L;\theta_L) + f_L\Delta(q_L) \tag{3.16}$$

and note that $u_H - s_H^* > 0 \iff \Gamma(f_L) > 0$.

Differentiating this function,

$$\Gamma'(f_L) = -s_H^* + s(q_L;\theta_L) + \Delta(q_L) + f_L\Delta'(q_L)[1 + \Phi(q_L)]\frac{\partial q_L}{\partial f_L} \tag{3.17}$$

From the F.O.C.s,

$$s_H^* - u_H + f_L[-s_H^* + s(q_L;\theta_L) + u_H - u_L] = 0$$

So $-s_H^* + s(q_L;\theta_L) + u_H - u_L \gtrless 0 \iff u_H - s_H^* \gtrless 0$. Since for $\zeta = 0$, $u_H \geq s_H^*$ is not binding, we infer that this term is positive. Next, differentiating the optimality condition for q_L, we have

$$\frac{\partial q_L^\infty}{\partial f_L} = -\frac{1}{f_L}\frac{1 + \Phi(q_L)}{\Phi'(q_L)} > 0$$

This implies that $\Gamma'(f_L) > 0$. Finally, from the F.O.C.s as $f_L \to 1$, we have $q_L \to q^*$. Thus, $\lim_{f_L \to 1}\Gamma(f_L) = -s_H^* + s_L^* + \Delta(q^*) > 0$. Also, note that $\Gamma(f_L) = -s_H^* + f_H s_H^* + f_L s(q_L;\theta_L) + f_L\Delta(q_L)$ so $\lim_{f_L \to 0}\Gamma(f_L) = 0$.

In other words, for sufficiently large f_L, the interim efficient allocation has cross-subsidisation. Furthermore, there exists a threshold level of f_L below which, $u_H = s_H^*$ and so there is no cross-subsidisation. We can reframe these findings as a threshold value of $\frac{f_H}{f_L}$ as follows.

Define $\phi^0 \equiv \Phi(q_L^0)$ where $s(q_L^0;\theta_L) - s_H^* + \Delta(q_L^0) = 0$ and recall $\Phi(q_L^\infty) = \frac{f_H}{f_L}$. The interim efficient allocation is such that

$$u_H^0, u_L^0, q_L^0 \quad s.t. \quad \begin{cases} u_H^0 = s_H^*,\ u_L^0 = s(q_L^0;\theta_L) & if \quad \frac{f_H}{f_L} \geq \phi^0 \\ u_H^0 > s_H^*,\ u_L^0 < s(q_L^\infty;\theta_L) & if \quad \frac{f_H}{f_L} < \phi^0 \end{cases}$$

If the LCS allocation is an equilibrium, then it must be interim efficient. Otherwise, there exists another menu which yields greater total surplus and which can be used to construct a profitable deviation. If an allocation is interim efficient, then it must be the least cost separating allocation.

A The duopoly insurance problem

Denote the monopoly optimum by

$$\begin{aligned} \{(u_i^m, q_i^m)\}_{i=H,L} \in \max_{\{(u_i,q_i)\}_{i=H,l}} \quad & f_H H(u_H; u_H^1, \tau)[s(q_H;\theta_H) - u_H] + f_L H(u_L; u_L^1, \tau)[s(q_L;\theta_L) - u_L] && (\mathcal{P}^d) \\ s.t. \quad & u_H \geq u_L + \Delta(q_L) && (IC_H) \\ & u_L \geq u_H - \Delta(q_H) && (IC_L) \\ & u_L \geq 0,\ q_L \geq 0 \end{aligned}$$

A.1 Characterisation

Monotonicity in coverage

First, note combining (IC_H) and (IC_L) yields $\Delta(q_H) \geq u_H - u_L \geq \Delta(q_L)$ so $q_H \geq q_L$.

Furthermore, this also means the constraint $q_H \geq 0$ is redundant. Also, since $\Delta(q_L) \geq 0$ for all $q_L \geq 0$, any allocation which satisfies (IC_H) has $u_H \geq u_L$. Hence the constraint $u_H \geq 0$ is redundant.

IC constraints cannot be all simultaneously binding

We first establish the following well-known result that both incentive compatibility constraints cannot be simultaneously binding.

Lemma 17. *A solution to the monopoly problem cannot simultaneously have $\lambda_H > 0$ and $\lambda_L > 0$.*

Proof. Suppose $\lambda_H > 0$ and $\lambda_L > 0$. This implies IC_H and IC_L hold with equality which, in turn, implies $\Delta(q_L) = u_H - u_L = \Delta(q_H)$ and $q_H = q_L$. At the same time, the derivative of the objective with respect to q_H reads $f_H H(u_H; u_H^1, \tau) s_q(q_H; \theta_H) = -\lambda_L \Delta'(q_H) < 0$ so $q_H > q^*$. First take $\varphi > 0$. Then we have an immediate contradiction since the binding constraint implies $q_L = 0 < q^* < q_H$. Next take $\varphi = 0$. Then the derivative of the objective with respect to q_L reads $f_L H(u_L; u_L^1, \tau) s_q(q_L; \theta_L) = \lambda_H \Delta'(q_L) > 0$ and hence $q_L < q^*$. Thus, $\lambda_H > 0$ and $\lambda_L > 0$ implies $q_H > q^* > q_L$, a contradiction. □

IC_H is binding for interior solutions

The next lemma establishes the familiar result that *interior solutions* of the insurer's problem must have the IC_H constraint binding.

Lemma 18. *Any interior solution to the duopoly insurance problem must be such that $\lambda_H > 0$.*

Proof. Any interior solution to the sub-problems must satisfy the necessary F.O.C.s

$$f_H \frac{\partial H(u_H; u_H^1, \tau)}{\partial u}[s(q_H; \theta_H) - u_H] - f_H H(u_H; u_H^1, \tau) + \lambda_H - \lambda_L = 0 \tag{3.18}$$

$$f_L \frac{\partial H(u_L; u_L^1, \tau)}{\partial u}[s(q_L; \theta_L) - u_L] - f_L H(u_L; u_L^1, \tau) - \lambda_H + \lambda_L = 0 \tag{3.19}$$

$$f_H H(u_H; u_H^1, \tau) s_q(q_H; \theta_H) + \lambda_L \Delta'(q_H) = 0 \tag{3.20}$$

$$f_L H(u_L; u_L^1, \tau) s_q(q_L; \theta_L) - \lambda_H \Delta'(q_L) + \varphi = 0 \tag{3.21}$$

$$\lambda_H[u_H - u_L - \Delta(q_L)] = 0,\ \lambda_H \geq 0,\ u_H - u_L - \Delta(q_L) \geq 0$$

$$\lambda_L[u_L - u_H + \Delta(q_H)] = 0,\ \lambda_L \geq 0,\ u_L - u_H + \Delta(q_H) \geq 0$$

$$\varphi q_L = 0,\ \varphi \geq 0,\ q_L \geq 0$$

Suppose, for a contradiction, that $\lambda_H = 0$ at the solution to this system. Then (3.18) and

(3.19) imply

$$[s(q_H;\theta_H)-u_H]-[s(q_L;\theta_L)-u_L]\geq\Big[\frac{1}{H(u_H;\tau)}\frac{\partial H(u_H;\tau)}{\partial u}\Big]^{-1}-\Big[\frac{1}{H(u_L;\tau)}\frac{\partial H(u_L,\tau)}{\partial u}\Big]^{-1}$$

Next, note that when $u_L > 0$, (3.21) implies we have $q_L = q^*$ while (3.20) implies $q_H \geq q^*$. Adding $s(q_H;\theta_H)$ and subtracting $s(q_L;\theta_L)$ to both sides of (IC_L), we have

$$\begin{aligned}[s(q_H;\theta_H)-u_H]-[s(q^*;\theta_L)-u_L] =& s(q_H;\theta_H)-s(q^*;\theta_L)-\Delta(q^*)\\ =& -c(q_H;\theta_H)+c(q^*;\theta_L)\leq -c(q^*;\theta_H)+c(q^*;\theta_L)<0\end{aligned}$$

We thus obtain our desired contradiction. □

Lemmas 17 and 18 imply that any interior solution must be such that $\lambda_H > 0$ and $\lambda_L = 0$. This, in turn, implies $q_H = q^*$ and $u_H = u_L + \Delta(q_L)$.

Necessary conditions

A solution to the problem must satisfy the following necessary F.O.C.s

$$f_H\frac{\partial H(u_H;u_H^1,\tau)}{\partial u}[s_H^*-u_H]-f_H H(u_H;u_H^1,\tau)+\lambda_H=0 \tag{3.22}$$

$$f_L\frac{\partial H(u_L;u_L^1,\tau)}{\partial u}[s(q_L;\theta_L)-u_L]-f_L H(u_L;u_L^1,\tau)-\lambda_H=0 \tag{3.23}$$

$$f_L H(u_L;u_L^1,\tau)s_q(q_L;\theta_L)-\lambda_H\Delta'(q_L)+\varphi=0 \tag{3.24}$$

$$u_H-u_L-\Delta(q_L)=0 \tag{3.25}$$

$$\varphi q_L=0,\ \varphi\geq 0,\ q_L\geq 0$$

Suppose $q_L = 0$, then the left-hand side of (3.23) reads $-f_L\frac{\partial H(u_L;u_L^1,\tau)}{\partial u}u_L - f_L H(u_L;u_L^1,\tau) - \lambda_H < 0$, thereby violating the necessary first-order condition. Therefore, we must have $q_L > 0$ and $\varphi = 0$ at any interior solution to the duopolist's problem.

A symmetric equilibrium is thus given by

$$f_H\frac{1}{\tau}[s_H^*-u_H]-f_H+f_L\Phi(q_L)=0 \tag{3.26}$$

$$f_L[s(q_L;\theta_L)-u_L]-f_H[s_H^*-u_H]-\tau=0 \tag{3.27}$$

$$u_H-u_L-\Delta(q_L)=0$$

where $\Phi(q) \equiv \frac{s_q(q;\theta_L)}{\Delta'(q)}$.

A.2 Sufficiency for symmetric equilibrium

We can show that if $\frac{f_H}{f_L} \geq \left[\frac{1}{2}\Phi(0)\right]^2$, then the objective is strictly concave in (u_H, u_L, q_L) and hence F.O.C.s (3.22) to (3.25) are necessary and sufficient for an interior solution to the duopolist's problem.

Let $\bar{C} \equiv \{(\bar{u}_H, \bar{q}_H), (\bar{u}_L, \bar{q}_L)\}$ denote the interior symmetric allocation given by (3.22) to (3.25). For this allocation to indeed be an equilibrium, it remains to rule out profitable market cornering strategies.

Cornering the low-risk segment

Consider a deviation from $\bar{C}$ to a menu which corners the low-risk segment. i.e. a menu which attracts all low-risk individuals to the Insurer-0. Let this deviation be denoted by $\hat{C} \equiv \{(\hat{u}_H, \hat{q}_H), (\hat{u}_L, \hat{q}_L)\}$. The profit-maximising menu that corners the low-risk segment solves

$$\max_{\{(u_i,q_i)\}_{i=H,L}} \quad f_H H(u_H; \bar{u}_H, \tau)[s(q_H;\theta_H) - u_H] + f_L[s(q_L;\theta_L) - u_L] \tag{$\hat{\mathcal{P}}_l$}$$

$$\text{s.t.} \quad u_H \geq u_L + \Delta(q_L) \tag{IC_H}$$

$$u_L \geq u_H - \Delta(q_H) \tag{IC_L}$$

$$u_L \geq \bar{u}_L + \tau, \; q_L \geq 0$$

The same arguments which imply that both IC constraints cannot be simultaneously binding continue to apply. Let ζ denote the Lagrange multiplier associated with the strict market cornering condition, $\bar{u}_L \geq u_L + \tau$. Any interior solution to the sub-problems must satisfy the necessary F.O.C.s

$$f_H \frac{\partial H(\hat{u}_H;\bar{u}_H,\tau)}{\partial u}[s(\hat{q}_H;\theta_H)-\hat{u}_H]-f_H H(\hat{u}_H;\bar{u}_H,\tau)+\lambda_H-\lambda_L=0 \tag{3.28}$$

$$-f_L-\lambda_H+\lambda_L+\zeta=0 \tag{3.29}$$

$$f_H H(\hat{u}_H;\bar{u}_H,\tau)s_q(\hat{q}_H;\theta_H)+\lambda_L\Delta'(\hat{q}_H)=0 \tag{3.30}$$

$$f_L s_q(\hat{q}_L;\theta_L)-\lambda_H\Delta'(\hat{q}_L)+\varphi=0 \tag{3.31}$$

$$\lambda_H[\hat{u}_H-\hat{u}_L-\Delta(\hat{q}_L)]=0,\ \lambda_H\geq 0,\ \hat{u}_H-\hat{u}_L-\Delta(\hat{q}_L)\geq 0$$

$$\lambda_L[\hat{u}_L-\hat{u}_H+\Delta(\hat{q}_H)]=0,\ \lambda_L\geq 0,\ \hat{u}_L-\hat{u}_H+\Delta(\hat{q}_H)\geq 0$$

$$\varphi\hat{q}_L=0,\ \varphi\geq 0,\ \hat{q}_L\geq 0$$

$$\zeta[\hat{u}_L-\bar{u}_L-\tau]=0,\ \zeta\geq 0,\ \hat{u}_L\geq\bar{u}_L+\tau$$

From (3.29), we must have either $\lambda_L>0$ or $\zeta>0$.

If $\zeta>0$, then $\hat{u}_L=u_L+\tau$ is binding and a local deviation to $\tilde{u}_L<u_L+\tau$ will increase total expected profit. However, this means cornering the entire low-risk segment is dominated by an interior strategy, i.e. a menu which attracts strictly less than all low-risk individuals. Therefore, for a profitable low-risk cornering deviation, we must have $\zeta=0$.

When $\zeta=0$, we must have $\lambda_L>0$ and hence $\lambda_H=0$. This, in turn, implies $q_H>q^*$ and $q_L=q^*$ with $\varphi=0$ and

$$\hat{u}_H=\hat{u}_L+\Delta(\hat{q}_H)\geq\tau+\bar{u}_L+\Delta(\hat{q}_H)\geq\tau+\bar{u}_H-\Delta(q^*)+\Delta(\hat{q}_H) \tag{3.32}$$

The first equality is the binding IC_L constraint for the cornering menu, the first inequality arises from the condition for cornering low-risks and the second inequality is from the non-binding IC_L constraint for the symmetric interior menu. Since $\hat{q}_H>q^*$ the sequence of inequalities above imply $\hat{u}_H-\bar{u}_H>\tau$. This, in turn, means the low-risk cornering menu must also corner the high-risk segment. However, substituting $\hat{u}_H-\bar{u}_H>\tau$ into F.O.C. (3.28) yields $-\lambda_L=0$, a contradiction. Because any incentive compatible deviation which corners the low-risk segment must satisfy the necessary F.O.C.'s and we have shown that such an allocation does not exist, we conclude there does not exist a profitable deviation from an interior symmetric candidate to a menu that corners the low-risk segment.

Cornering the high-risk segment

Consider a deviation from $\bar{C} \equiv \{(\bar{u}_H, \bar{q}_H), (\bar{u}_L, \bar{q}_L)\}$ to a menu which corners the high-risk segment. i.e. a menu which attracts all high-risk individuals to the Insurer-0. Let this deviation be denoted by $\hat{C} \equiv \{(\hat{u}_H, \hat{q}_H), (\hat{u}_L, \hat{q}_L)\}$. $\hat{C}$ must be such that $\hat{u}_H \geq u_H + \tau$. Given that the deviation must still be incentive compatible, we have $\hat{u}_H \geq \hat{u}_L + \Delta(\hat{q}_L) \geq u_L + \Delta(\hat{q}_L)$.

The profit-maximising menu that corners the high-risk segment solves

$$
\begin{aligned}
\max_{\{(\bar{u}_i, \bar{q}_i)\}_{i=H,L}} \quad & f_H[s(\bar{q}_H; \theta_H) - \bar{u}_H] + f_L H(\bar{u}_L; u_L, \tau)[s(\bar{q}_L; \theta_L) - \bar{u}_L] && (\hat{\mathcal{P}}_l) \\
s.t. \quad & \bar{u}_H \geq \bar{u}_L + \Delta(\bar{q}_L) && (IC_H) \\
& \bar{u}_L \geq \bar{u}_H - \Delta(\bar{q}_H) && (IC_L) \\
& \bar{u}_H \geq u_H + \tau, \ \bar{q}_L \geq 0 &&
\end{aligned}
$$

The same arguments which imply that both IC constraints cannot be simultaneously binding continue to apply. Let ζ denote the Lagrange multiplier associated with the strict market cornering condition, $\bar{u}_H \geq u_H + \tau$. Any interior solution to the sub-problems must satisfy the necessary F.O.C.s

$$
\begin{aligned}
& -f_H + \lambda_H - \lambda_L + \zeta = 0 && (3.33) \\
& f_L \frac{\partial H(u_L; u_L^1, \tau)}{\partial u}[s(q_L; \theta_L) - u_L] - f_L H(u_L; u_L^1, \tau) - \lambda_H + \lambda_L = 0 && (3.34) \\
& f_H s_q(q_H; \theta_H) + \lambda_L \Delta'(q_H) = 0 && (3.35) \\
& f_L H(u_L; u_L^1, \tau) s_q(q_L; \theta_L) - \lambda_H \Delta'(q_L) + \varphi = 0 && (3.36) \\
& \lambda_H[u_H - u_L - \Delta(q_L)] = 0, \ \lambda_H \geq 0, \ u_H - u_L - \Delta(q_L) \geq 0 && \\
& \lambda_L[u_L - u_H + \Delta(q_H)] = 0, \ \lambda_L \geq 0, \ u_L - u_H + \Delta(q_H) \geq 0 && \\
& \varphi q_L = 0, \ \varphi \geq 0, \ q_L \geq 0 &&
\end{aligned}
$$

Suppose $\lambda_H = 0$, then for F.O.C. (3.33) to hold, we must have $\zeta > 0$. However, this implies that total expected profit can be increased by offering $u_H < u_H + \tau$. This means cornering the high-risk segment is not globally optimal.

Therefore, we must have $\hat{\lambda}_H > 0$ and $\hat{\lambda}_L = 0$. This, in turn, implies $\hat{q}_H = q^*$ and $\hat{u}_H = \hat{u}_L + \Delta(\hat{q}_L)$. Suppose $\varphi > 0$ and hence $q_L = 0$. This violates (3.34) since the left-hand side would be strictly negative. Hence we must have $\hat{\varphi} = 0$ and $0 < \hat{q}_L < q^*$. Substituting these conditions, the F.O.C.s reduce to

$$-f_H + \lambda_H + \zeta = 0 \tag{3.37}$$

$$f_L[s(q_L;\theta_L) - u_L] - f_L[\tau + \hat{u}_L - u_L] - 2\tau\lambda_H = 0 \tag{3.38}$$

$$f_L[\tau + \hat{u}_L - u_L]\Phi(q_L) - 2\tau\lambda_H = 0 \tag{3.39}$$

Suppose $\zeta = 0$. Substituting (3.37) and (3.39), we have

$$-\frac{f_H}{f_L} + H(\hat{u}_L; u_L, \tau)\Phi(\hat{q}_L) + \frac{1}{f_L}\zeta = 0 \tag{3.40}$$

Since $H(u,\tau) \leq 1$ and $\Phi(q) \leq \Phi(0)$, if $\frac{f_H}{f_L} > \Phi(0)$, then cornering the high-risk segment is dominated by an interior strategy for all $\tau \geq 0$.

Suppose $\frac{f_H}{f_L} \leq \Phi(0)$. Then using the fact that IC_H is binding in both symmetric and cornering strategies we have $\zeta > 0$ if and only if

$$-\frac{f_H}{f_L} + \frac{1}{2\tau}\Big[\tau + (u_H + \tau) - u_H - \Delta(\hat{q}_L) + \Delta(q_L)\Big]\Phi(\hat{q}_L) < 0 \iff \frac{f_H}{f_L} > \Big[1 - \frac{\Delta(\hat{q}_L) - \Delta(q_L)}{2\tau}\Big]\Phi(\hat{q}_L)$$

We can show that $\hat{q}_L$ is strictly increasing in τ. This means the term on the right-hand side of the inequality approaches zero as τ increases without bound. Conversely, because we must necessarily have $\hat{q}_L \geq q_L$, the right-hand side is bounded above as τ approaches zero.

Therefore, for each τ there is a threshold risk distribution, $\underline{\phi} \equiv \Phi(\hat{q}_L) - [\Delta(\hat{q}_L) - \Delta(q_L)]\Phi(\hat{q}_L)/2\tau$ such that any high-risk market cornering strategy is dominated by the symmetric interior candidate equilibrium whenever $f_H/f_L \geq \underline{\phi}$.

Symmetrically, for each given f_H/f_L, there exists a $\underline{\tau}$ implicitly defined by $\frac{f_H}{f_L} = \Phi(\hat{q}_L(\underline{\tau})) - [\Delta(\hat{q}_L(\underline{\tau})) - \Delta(q_L(\underline{\tau}))]\Phi(\hat{q}_L(\underline{\tau}))/2\underline{\tau}$ such that any high-risk market cornering strategy is dominated by the symmetric interior candidate equilibrium whenever $\tau \geq \underline{\tau}$.

A.3 Comparative statics

The second partial derivatives of the Lagrangean of the problem are

$$\begin{aligned}
\frac{\partial^2 \mathcal{L}}{\partial u_H^2} =& - f_H \frac{1}{\tau} \\
\frac{\partial^2 \mathcal{L}}{\partial u_L^2} =& - f_L \frac{1}{\tau} \\
\frac{\partial^2 \mathcal{L}}{\partial q_L^2} =& f_L \Big[\frac{1}{2} + \frac{u_L - u_L^{-k}}{2\tau}\Big] s_{qq}(q_L;\theta_L) - \lambda \Delta''(q_L) \\
\frac{\partial^2 \mathcal{L}}{\partial q_L \partial u_L} =& f_L \frac{1}{2\tau} s_q(q_L;\theta_L) \\
\frac{\partial^2 \mathcal{L}}{\partial \tau \partial u_H} =& - \frac{1}{\tau^2} f_H \frac{1}{2}[s_H^* - u_H] + \frac{1}{\tau^2} f_H \frac{u_H - u_H^{-k}}{2} \\
\frac{\partial^2 \mathcal{L}}{\partial \tau \partial u_L} =& - \frac{1}{\tau^2} f_L \frac{1}{2}[s(q_L;\theta_L) - u_L] + \frac{1}{\tau^2} f_L \frac{u_L - u_L^{-k}}{2} \\
\frac{\partial^2 \mathcal{L}}{\partial \tau \partial q_L} =& - \frac{1}{\tau^2} f_L \frac{u_L - u_L^{-k}}{2} s_q(q_L;\theta_L)
\end{aligned}$$

So $\mathcal{L}_{HH} < 0$, $\mathcal{L}_{LL} < 0$, $\mathcal{L}_{qq} < 0$, $\mathcal{L}_{Lq} > 0$ and $\mathcal{L}_{H\tau} \leq 0$, $\mathcal{L}_{L\tau} < 0$, $\mathcal{L}_{q\tau} \leq 0$

Applying Cramer's Rule,

$$\begin{aligned}
|B| \frac{\partial u_H}{\partial \tau} =& \mathcal{L}_{qq}[\mathcal{L}_{H\tau} + \mathcal{L}_{L\tau}] - \mathcal{L}_{Lq}\mathcal{L}_{q\tau} + (\Delta')^2 \mathcal{L}_{LL}\mathcal{L}_{H\tau} - 2\Delta' \mathcal{L}_{Lq}\mathcal{L}_{H\tau} \\
& + \Delta' \mathcal{L}_{LL}\mathcal{L}_{q\tau} - \Delta' \mathcal{L}_{Lq}\mathcal{L}_{L\tau} \\
|B| \frac{\partial u_L}{\partial \tau} =& \mathcal{L}_{qq}[\mathcal{L}_{H\tau} + \mathcal{L}_{L\tau}] - \mathcal{L}_{Lq}\mathcal{L}_{q\tau} - \Delta' \mathcal{L}_{Lq}\mathcal{L}_{H\tau} \\
& + (\Delta')^2 \mathcal{L}_{HH}\mathcal{L}_{L\tau} - \Delta' \mathcal{L}_{HH}\mathcal{L}_{q\tau} \\
|B| \frac{\partial q_L}{\partial \tau} =& [\mathcal{L}_{HH} + \mathcal{L}_{LL}]\mathcal{L}_{q\tau} - \mathcal{L}_{Lq}[\mathcal{L}_{H\tau} + \mathcal{L}_{L\tau}] \\
& + \Delta' \mathcal{L}_{LL}\mathcal{L}_{H\tau} - \Delta' \mathcal{L}_{HH}\mathcal{L}_{L\tau}
\end{aligned}$$

where $|B|$ is the determinant of the Hessian of $\mathcal{L}$.

In a symmetric equilibrium, $u_i = u_i^{-k}$, so $\mathcal{L}_{q\tau} = 0$ and $\mathcal{L}_{H\tau} + \mathcal{L}_{L\tau} = -\frac{1}{\tau}\Big[f_H \frac{1}{2\tau}[s_H^* - u_H] +$

$f_L \frac{1}{2\tau}[s(q_L;\theta_L) - u_L]\Big] = -\frac{1}{2\tau}$ and we have

$$
\begin{aligned}
|B|\frac{\partial u_H}{\partial \tau} &= \mathcal{L}_{qq}[\mathcal{L}_{H\tau} + \mathcal{L}_{L\tau}] + (\Delta')^2 \mathcal{L}_{LL}\mathcal{L}_{H\tau} - 2\Delta' \mathcal{L}_{Lq}\mathcal{L}_{H\tau} - \Delta' \mathcal{L}_{Lq}\mathcal{L}_{L\tau} > 0 \\
|B|\frac{\partial u_L}{\partial \tau} &= \mathcal{L}_{qq}[\mathcal{L}_{H\tau} + \mathcal{L}_{L\tau}] - \Delta' \mathcal{L}_{Lq}\mathcal{L}_{H\tau} + (\Delta')^2 \mathcal{L}_{HH}\mathcal{L}_{L\tau} > 0 \\
|B|\frac{\partial q_L}{\partial \tau} &= \frac{1}{4\tau^2} f_L s_q - \Delta' f_L \frac{1}{2\tau^2}\Phi = -\Delta' f_L \frac{1}{2\tau^2}\frac{1}{2}\Phi < 0
\end{aligned}
$$

Since $|B| < 0$ whenever the Lagrangean is concave, we have $\frac{\partial u_H}{\partial \tau} < 0$, $\frac{\partial u_L}{\partial \tau} < 0$ and $\frac{\partial q_L}{\partial \tau} > 0$.

Bibliography

Ameriks, J., Briggs, J., Caplin, A., Shapiro, M. D., and Tonetti, C. (2016). The long-term-care insurance puzzle: Modeling and measurement. Technical report, National Bureau of Economic Research.

Ameriks, J., Briggs, J., Caplin, A., Shapiro, M. D., and Tonetti, C. (2020). Long-term-care utility and late-in-life saving. *Journal of Political Economy*, 128(6):2375–2451.

Bénabou, R. and Tirole, J. (2016). Bonus culture: Competitive pay, screening, and multitasking. *Journal of Political Economy*, 124(2):305–370.

Bester, H. et al. (1985). Screening vs. rationing in credit markets with imperfect information. *American economic review*, 75(4):850–855.

Bijlsma, M., Boone, J., and Zwart, G. (2014). Competition leverage: how the demand side affects optimal risk adjustment. *The RAND Journal of Economics*, 45(4):792–815.

Boyer, M., De Donder, P., Fluet, C., Leroux, M.-L., and Michaud, P.-C. (2020). Long-term care insurance: Information frictions and selection. *American Economic Journal: Economic Policy*.

Braun, R. A., Kopecky, K. A., and Koreshkova, T. (2019). Old, frail, and uninsured: Accounting for features of the us long-term care insurance market. *Econometrica*, 87(3):981–1019.

Brown, J. R. and Finkelstein, A. (2008). The interaction of public and private insurance: Medicaid and the long-term care insurance market. *American Economic Review*, 98(3):1083–1102.

Brown, J. R. and Finkelstein, A. (2011). Insuring long-term care in the united states. *Journal of Economic Perspectives*, 25(4):119–42.

Chade, H. and Schlee, E. (2012). Optimal insurance with adverse selection. *Theoretical Economics*, 7:571–607.

Chade, H. and Schlee, E. E. (2020). Insurance as a lemons market: Coverage denials and pooling. *Journal of Economic Theory*, 189:105085.

Chetty, R. (2006). A general formula for the optimal level of social insurance. *Journal of Public Economics*, 90(10-11):1879–1901.

Chetty, R. and Szeidl, A. (2007). Consumption commitments and risk preferences. *The Quarterly Journal of Economics*, 122(2):831–877.

Dafny, L., Duggan, M., and Ramanarayanan, S. (2012). Paying a premium on your premium? consolidation in the us health insurance industry. *The American Economic Review*, 102(2):1161–1185.

Dafny, L. S. (2010). Are health insurance markets competitive? *The American Economic Review*, 100(4):1399–1431.

Davidoff, T. (2010). Home equity commitment and long-term care insurance demand. *Journal of Public Economics*, 94(1-2):44–49.

De Nardi, M., French, E., and Jones, J. B. (2010). Why do the elderly save? the role of medical expenses. *Journal of political economy*, 118(1):39–75.

Dobkin, C., Finkelstein, A., Kluender, R., and Notowidigdo, M. J. (2018). The economic consequences of hospital admissions. *American Economic Review*, 108(2):308–52.

Fadlon, I. and Nielsen, T. H. (2019). Household labor supply and the gains from social insurance. *Journal of public economics*, 171:18–28.

Finkelstein, A., Luttmer, E. F., and Notowidigdo, M. J. (2013). What good is wealth without health? the effect of health on the marginal utility of consumption. *Journal of the European Economic Association*, 11(suppl_1):221–258.

Finkelstein, A. and McGarry, K. (2006). Multiple dimensions of private information: evidence from the long-term care insurance market. *American Economic Review*, 96(4):938–958.

French, E., Jones, J. B., Kelly, E., and McCauley, J. (2018). End-of-life medical expenses.

Friedberg, L., Hou, W., Sun, W., Webb, A., and Li, Z. (2014). New evidence on the risk of requiring long-term care. *Boston College Center for Retirement Research WP*, 12.

Ganong, P. and Noel, P. (2019). Consumer spending during unemployment: Positive and normative implications. *American Economic Review*, 109(7):2383–2424.

Gruber, J. (1997). The consumption smoothing benefits of unemployment insurance. *The American Economic Review*, 87(1):192.

Hendren, N. (2013). Private information and insurance rejections. *Econometrica*, 81(5):1713–1762.

Hendren, N. (2017). Knowledge of future job loss and implications for unemployment insurance. *American Economic Review*, 107(7):1778–1823.

Hurd, M. D., Michaud, P.-C., and Rohwedder, S. (2014). The lifetime risk of nursing home use. In *Discoveries in the Economics of Aging*, pages 81–109. University of Chicago Press.

Jack, W. (2006). Optimal risk adjustment with adverse selection and spatial competition. *Journal of Health Economics*, 25(5):908–926.

Koijen, R. S., Van Nieuwerburgh, S., and Yogo, M. (2016). Health and mortality delta: Assessing the welfare cost of household insurance choice. *The Journal of Finance*, 71(2):957–1010.

Kolsrud, J., Landais, C., Nilsson, P., and Spinnewijn, J. (2018). The optimal timing of unemployment benefits: Theory and evidence from sweden. *American Economic Review*, 108(4-5):985–1033.

Kolsrud, J., Landais, C., and Spinnewijn, J. (2019). The value of registry data for consumption analysis: An application to health shocks. *Journal of Public Economics*, page 104088.

Landais, C. and Spinnewijn, J. (2019). The value of unemployment insurance.

Lockwood, L. M. (2018). Incidental bequests and the choice to self-insure late-life risks. *American Economic Review*, 108(9):2513–50.

Meyer, B. D. and Mok, W. K. (2019). Disability, earnings, income and consumption. *Journal of Public Economics*, 171:51–69.

Olivella, P. and Vera-Hernández, M. (2007). Competition among differentiated health plans under adverse selection. *Journal of Health Economics*, 26(2):233–250.

Rochet, J.-C. and Stole, L. A. (2002). Nonlinear pricing with random participation. *The Review of Economic Studies*, 69(1):277–311.

Rothschild, M. and Stiglitz, J. (1976). Equilibrium in competitive insurance markets: An essay on the economics of imperfect information. *The Quarterly Journal of Economics*, 90(4):629–649.

Shen, J., Yang, H., and Ye, L. (2016). Competitive nonlinear pricing and contract variety. *The Journal of Industrial Economics*, 64(1):64–108.

Stiglitz, J. E. (1977). Monopoly, non-linear pricing and imperfect information: The insurance market. *Review of Economic Studies*, 44(3):407–430.

Villas-Boas, J. M. and Schmidt-Mohr, U. (1999). Oligopoly with asymmetric information: differentiation in credit markets. *The RAND Journal of Economics*, pages 375–396.

Wilson, C. (1977). A model of insurance markets with incomplete information. *Journal of Economic Theory*, 16:167–207.

Yang, H. and Ye, L. (2008). Nonlinear pricing, market coverage, and competition. *Theoretical Economics*, pages 123–153.