NBER WORKING PAPER SERIES

MORAL HAZARD UNDER LIQUIDITY CONSTRAINTS

Keith Marzilli Ericson
Johannes G. Jaspersen
Justin R. Sydnor

Working Paper 33648
http://www.nber.org/papers/w33648

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2025

We acknowledge funding from the Boston University Questrom School of Business. None of the authors has a conflict of interest with this work. We thank Stuart Craig, Cecilia Diaz-Campo, Amy Finkelstein, Nathaniel Hendren, Wanda Mimra, Philip Mulder, Joseph Newhouse, and Dan Sacks, as well as participants at the 2024 annual meeting of the American Society of Health Economists, the University of Wisconsin, Harvard Health Care and Policy, and the CEAR/ MRIC Behavioral Insurance Workshop for helpful comments. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

Moral Hazard under Liquidity Constraints
Keith Marzilli Ericson, Johannes G. Jaspersen, and Justin R. Sydnor
NBER Working Paper No. 33648
April 2025
JEL No. D15, D80, H30, I13

ABSTRACT

Spending induced by health insurance is often called moral hazard and definitionally assumed to be inefficient. We adapt standard models and show that for those living "hand-to-mouth", the financing benefits of insurance cause a portion of moral hazard to be efficient. Although insurance's price distortions also create some inefficient spending, the net welfare impacts of moral hazard can be positive. We present an intuitive graphical framework and formal results to distinguish moral hazard's efficient and inefficient components. Simulations show economically significant net benefits of moral hazard in many cases. Our framework also provides a new way of modeling the "income effect'" induced by insurance, and distinguishes it from the "liquidity effect". While both can lead to efficient moral hazard, moral-hazard benefits from the "liquidity effect" are often substantially larger. We use our framework to revisit prior estimates of Medicaid's value from the Oregon Health Insurance Experiment. For individuals with minimal liquidity, Medicaid's value is more than twice prior estimates.

Keith Marzilli Ericson
Boston University
Questrom School of Business
595 Commonwealth Avenue
Boston, MA 02215
and NBER
kericson@bu.edu

Justin R. Sydnor
Wisconsin School of Business,
ASRMI Department
University of Wisconsin-Madison
975 University Avenue, Room 5287
Madison, WI 53726
and NBER
jsydnor@bus.wisc.edu

Johannes G. Jaspersen
Ludwig-Maximilians-Universität of Munich
Schackstr. 4
80539 Munich
Germany
jaspersen@lmu.de

1 Introduction

The change in spending resulting from health insurance coverage is typically referred to as moral hazard.[1] This empirical definition—a change in spending—is often paired with a welfare interpretation: moral hazard is considered "excess" spending that individuals value less than its full cost due to a distortion in incentives (Pauly, 1968). The dominant modeling frameworks used in the analysis of the value of health insurance imply that moral hazard lowers welfare (Einav et al., 2013; Finkelstein et al., 2019; Marone and Sabety, 2022).

We revisit this interpretation of moral hazard and build on insights from De Meza (1983), Nyman (1999b), and Chetty (2008) that for liquidity constrained individuals, moral hazard can actually be efficient and welfare increasing. We emphasize how insurance provides both a risk-protection benefit and access to financing: while insured expenses are often concentrated in a particular consumption period, the cost of insurance via premiums is typically spread across multiple consumption periods. An important feature of health insurance is that it can smooth expenditures across time, not merely across states (e.g. Ericson and Sydnor, 2018; Gross et al., 2022), and that this can quantitatively and qualitatively change patterns of insurance demand and spending.

Accounting for liquidity constraints is important, particularly for evaluating expansions of insurance to low-income individuals. Many individuals face liquidity constraints (either the inability to borrow or high interest rates) and have high marginal-propensities to consume out of income shocks (Parker et al., 2013). As a result, macroeconomic models often treat a large share of individuals as living essentially hand-to-mouth, consuming all liquid resources in a given period (Kaplan et al., 2014; Lee and Maxted, 2023; Aguiar et al., 2024). Moreover, because of holdings in illiquid assets, hand-to-mouth behavior can describe even wealthy, optimizing individuals (Kaplan et al., 2014). Further, even if people ultimately have access to liquidity, if they have behavioral biases that lead them to act "as-if" they were hand-to-mouth (Olafsson and Pagel, 2018; Lee and Maxted, 2023), those perceived liquidity constraints will distort spending when uninsured.[2]

In this paper we analyze the welfare implications of ex-post moral hazard in light of liquidity constraints. Our setting is health insurance, which has been the focus of a substantial body of literature on moral hazard (Einav and Finkelstein, 2018), but the model is broadly applicable to various types of insurance. Standard models of ex-post moral hazard in (health) insurance are one-period models in which premiums, cost-sharing, and consumption all occur in the same consumption period. Our innovation, which closely follows the approach in Ericson and Sydnor (2018) is to introduce multiple underlying consumption periods within the insurance period. The cost of insurance via premiums is spread equally across all N consumption periods within the

[1] For instance, Einav and Finkelstein (2018) write "we follow decades of health insurance literature and use the term "moral hazard" to refer to the responsiveness of healthcare spending to insurance coverage". For a discussion of the history of the term and alternative definitions, see Rowell and Connelly (2012).

[2] In our model, we assume individuals are optimizing. However, new issues may arise if financial stress itself affects decision-making, as in Sergeyev et al. (2024).

insurance policy term. Uninsured medical expenses, however, are spread across $K \leq N$ periods. In this way we introduce a portable extension of the standard model, in which $K = N$ embeds the standard model that implicitly assumes perfect liquidity within the insurance-contract duration, while $K = 1$ is the full hand-to-mouth case, in which all uninsured expenses are borne in a single consumption period. This framework was previously used by Ericson and Sydnor (2018) to examine the value of insurance under liquidity constraints in the absence of moral hazard. The current paper explicitly focuses on moral hazard and the endogenous decision of medical spending.

We also innovate on existing discussions of moral hazard by explicitly defining the efficient level of medical spending for a given level of insurance coverage. We define efficient spending as the medical spending the individual would choose if they maximized ex-ante utility fully internalizing how their medical spending decision, when sick, would affect insurance premiums given the level of insurance. In other words, this represents the spending level they would select if they could contractually commit to a specific amount, paying for it through higher premiums. This efficient benchmark can then be compared to two alternative scenarios: the level of medical spending chosen when uninsured, and the level chosen when insured but taking the premiums for insurance as given (i.e., sunk). The chosen level of spending under insurance will be higher than the efficient level, because premiums are fixed for a contract period and insureds do not internalize the impact additional spending has on premiums. The innovation of this paper is to show how the uninsured level of spending can be below the efficient level for those with liquidity constraints.

Our framework allows us to present simple definitions decomposing moral hazard into both inefficient and *efficient* components. We develop intuitive graphical depictions of the model that allow us to illustrate the welfare impact of moral hazard and how it depends on liquidity constraints. This framework makes it clear how the welfare impacts of efficient and inefficient moral hazard can be quantified using the same data on medical spending chosen under different levels of insurance that have previously been used to evaluate the impact of moral hazard in empirical studies (e.g., Finkelstein et al., 2019).

Our theory first considers the standard intuition that moral hazard is inefficient in the absence of liquidity constraints. When the loss is certain to happen, that standard intuition holds: adding insurance increases expenditure, and this expenditure is not valued by individuals at its full cost. The level of spending that individual would contract on is invariant to the insurance level, meaning the privately chosen uninsured spending level was optimal. As a result, increasing insurance lowers welfare for the unconstrained.

For liquidity-constrained individuals, the results are quite different. The level of spending that an individual and insurer would jointly contract on now depends on the level of insurance, since insured spending (ultimately borne via premiums) is smoothed over consumption periods but cost-sharing hits in only a subset of periods. The efficient level of spending increases with the level of insurance. As a result, when spending increases with an expansion of insurance, this can increase rather than reduce welfare. Since insurance distorts prices, the individual will choose more than

the efficient level of spending when insured, but the benefits of increases in efficient moral hazard can outweigh the inefficient moral hazard.

We show in numerical simulations that the net welfare impacts of moral hazard can be significantly positive for individuals living hand-to-mouth. This updated welfare interpretation of moral hazard is relevant for setting the optimal level of cost-sharing. Since moral hazard can be net positive for those with liquidity constraints, the optimal level of insurance can also be much higher for them. In standard theory, when demand is more responsive to price– when there is more moral hazard– optimal cost-sharing should be higher (Zeckhauser, 1970; Pauly and Blavin, 2008). However, if environments with more moral hazard are settings where people have more liquidity constraints, the standard intuition could be backward.

Our framework also makes it possible to revisit an important insight from De Meza (1983) and Nyman (1999a): even in the absence of liquidity constraints, moral hazard is not entirely inefficient. These papers highlight that there is an income effect from insurance when the probability of illness is less than one. The main benefit of insurance comes from the ex-ante value of reducing spending risk into a sure premium. However, ex-post, when an individual is sick, they are also receiving a transfer from those who did not get sick and the optimal level of spending will be higher than when uninsured due to the income value of that transfer.[3] Our definitions of efficient and inefficient moral hazard and our graphical framework make it possible to intuitively visualize this "de Meza-Nyman income effect" and to quantify the welfare value of moral hazard given this income effect.[4]

We compare how liquidity constraints and the de Meza-Nyman income effect impact the value of moral hazard both theoretically and in simulations. Those with liquidity constraints get both a liquidity benefit and the de Meza-Nyman income effect from insurance, so the value of moral hazard is always more positive for those with liquidity constraints. In our simulations, for those without liquidity constraints the de Meza-Nyman income effect is generally not strong enough to make the net value of moral hazard positive.[5] For those living hand-to-mouth, however, the net value of moral hazard is significantly positive in many scenarios. Moreover, even for individuals with perfect liquidity, the de Meza-Nyman income effect becomes more important for very high levels of risk aversion and the net value of moral hazard becomes positive in some of our simulations

---

[3] This "de Meza-Nyman income effect" is different than the effect of an increase in permanent income on uninsured spending decisions. The "de Meza-Nyman income effect" depends on the probability of suffering illness and is stronger when that probability is low since that increases the share of spending that is transferred from non-sick states/individuals.

[4] While related, our definition of "efficient moral hazard" differs from the one presented in Nyman (1999a). Nyman defines it based on the difference between medical spending that would occur under a lump-sum transfer equal to the amount of insured spending under insurance and spending chosen when uninsured. We show in Appendix B that this lump sum amount is greater than what would induce the efficient spending level under our definition.

[5] This finding is naturally sensitive to the parameters of our simulations. Our simulations focus on modeling the overall value of health insurance over a range of parameters broadly similar to those in studies such as Finkelstein et al. (2019). It is quite plausible that for some types of coverage, especially protection against rare diseases for which there is little value of (or possibility of) treatment at lower levels of spending, that the net value of moral hazard can be positive, given the de Meza-Nyman income effect.

with low-probability events for the highly risk averse. While it may be tempting to conclude, then, that one can think of liquidity constraints simply as a case of very high effective risk aversion, we note that this is not the case because the de Meza-Nyman income effect relies on the likelihood of the sick state being low, while the liquidity benefit of insurance does not.

We use our framework to revisit Finkelstein et al. (2019)'s estimate of the value of Medicaid. Finkelstein et al. (2019) develop a framework for welfare analysis. Using data from the Oregon Health Insurance Experiment, they find that the willingness to pay for insurance by Medicaid recipients is well below the cost to the government. A key part of that analysis is a quantification of the value of the increased medical spending resulting from gaining insurance– the value of moral hazard. We revisit this analysis using our framework and analyze how the conclusions would be change if instead of assuming perfect liquidity during the insured policy year, we assumed that those receiving Medicaid were in hand-to-mouth situations and could not easily spread out uninsured spending across consumption periods. We find that the overall value of Medicaid would be more than twice as high if Medicaid members are living fully hand-to-mouth and experience their medical spending shocks in a single month. While the biggest part of the value of Medicaid comes from its risk and financing value without moral hazard, we estimate that the valuation of the moral hazard component of spending is substantially higher for the hand-to-mouth. We explore alternative assumptions about the concentration of medical spending across time. In our baseline model, the medical spending shock occurs in a single month, and only about the highest 25% of shocks are given payment plans to smooth across time. We also show, though, that if all uninsured medical spending instead is paid across three months, there are still very large increases in the value of Medicaid to those living hand-to-mouth. Ultimately, these estimates provide what we think are useful benchmarks to counterbalance the original estimates that implicitly assume no liquidity constraints within the insurance year. Given that there is ample evidence that low-income individuals often live in hand-to-mouth situations, this alternative estimate of the potential value of the program is important to consider. Moreover, this exercise shows how it is practically feasible to update the welfare estimates of health insurance under moral hazard to account for liquidity constraints.[6]

Our paper is closely related to Chetty (2008), which examines how unemployment durations respond to unemployment insurance benefits. Chetty's key insight is that while a portion of these responses are inefficient moral hazard driven by distorted incentives, part of the response can be efficient if it results from overcoming liquidity constraints. Our paper echoes this point and provides a new way of visualizing and developing intuition for it. The primary difference between our paper and Chetty (2008) lies in how we model and quantify the liquidity value of insurance.

[6] A related paper (Mukherjee et al., 2024) examines the extent to which Medicaid reduces individuals' consumption risk. It measures a subset of consumption expenses– "well-measured consumption", such food at home, rent, utilities, car expenses. It finds that these expenses do not respond much to the Medicaid expansion, and so estimate a small insurance value from Medicaid. Their results are consistent with many of the estimates from Finkelstein et al. (2019). However, they do not measure the smoothness of consumption across time, which is relevant in our model, and there are important elements of consumption not included in their measure.

Chetty derives a formula for the optimal benefit levels that depends on two different elasticities: a response to lump-sum cash and a response to incentives on the margin. A strength of this approach is that it relies on information from an empirical response, which itself can be consistent with various combinations of underlying liquidity constraints and preferences of optimizing agents. While this approach could be conceptually extended to analyze moral hazard in health insurance, Chetty and Finkelstein (2013) highlight that it requires developing quasi-experimental strategies for identifying lump-sum payments in health insurance that identify the liquidity effect. Although there is now some direct evidence that medical spending is liquidity sensitive (Gross et al., 2022), the liquidity effect of insurance has not been incorporated into analysis of moral hazard in health insurance. Our approach is an extension of the models currently used in the health insurance literature (e.g., Finkelstein et al., 2019). The benefit of this approach is that, as our empirical example shows, the same types of data and assumptions used in recent empirical analysis of moral hazard can be extended to incorporate liquidity constraints. A secondary point of differentiation between our work and Chetty (2008) is that Chetty's framework starts from the presumption that the individual is unemployed and abstracts from the risk-protection benefits of insurance, while we do not. Incorporating the uncertainty in whether the sick state will happen allows us to identify and distinguish the de Meza-Nyman income effect from the liquidity effect of insurance and to examine the magnitude of the net moral hazard value relative to the risk-protection and financing value of insurance. We see our work as providing a new way of visualizing and quantifying the key insight about the liquidity-effect component of moral hazard in a way that is closely linked to existing frameworks in health insurance.

Our paper is also related to a literature showing that people respond to the "spot" prices for medical care that they face at a given point in time, rather than merely the effective end-of-year price for that care (e.g., Aron-Dine et al., 2015; Brot-Goldberg et al., 2017). Liquidity constraints could help explain this phenomenon– the spot price rationally matters if people cannot shift money across time. However, the welfare implications of this response have not been previously explored.

2 Model

2.1 Layout

In our model, individuals live N periods[7] and have lifetime utility consisting of consumption and health components, $U = \sum_{t=1}^{N} \delta^{t-1} \left(u(c_t) + h(m_t, \theta_t) \right)$. We assume the exponential discount factor $\delta = 1$ throughout to highlight the role of liquidity constraints apart from time preference; with horizons less than a year, δ should be near 1 (Ericson and Laibson, 2019).

[7] This model and exposition build on Ericson and Sydnor (2018), adding the possibility of choosing medical spending as in Finkelstein et al. (2019). In contrast to the Finkelstein et al. (2019) model, we assume N consumption periods in a policy period rather than just one.

Consumption utility in each period t is given by $u(c_t)$, with $u(\cdot)$ continuous, $u' > 0$ and $u'' < 0$ and following the standard Inada conditions. Individuals receive total income y, paid in equal installments each period.

Utility from health is separable from consumption and is given each period by $h(m_t, \theta_t)$, where θ_t indexes the individual's health state and m_t indexes the level of medical spending in period t.

We make a set of simplifying assumptions for our our baseline model. We assume there are two health states: $\theta = sick$ and $\theta = healthy$. With probability $1 - \pi$ the individual is healthy for all N periods. With probability π the individual is sick in one of the N periods. We assume that the realization of the health state occurs at the beginning of the first period, so before choosing consumption the individual knows whether and when they will be sick.[8]

In the healthy state, health is unaffected by medical spending and $h(m, healthy)$ is normalized to be zero. In the sick state, medical spending initially generates a positive but decreasing marginal benefit. We suppress the θ argument when sick and write that $h'(m) > 0$ and $h''(m) < 0$ for all values of m smaller than $\bar{m}$. The function gives benefits of medical spending up to a point. Once $\bar{m}$ is reached, more medical spending has a negative effect on health such that $h'(m) < 0$ and $h''(m) < 0$. This is in line with the models of Einav et al. (2013) and Marone and Sabety (2022) and captures the idea that after reaching a certain health state additional medical spending is either useless (but still leads to detriments due the associated hassle costs) or actively harmful.

As a result of these assumptions, an individual's utility is simply $U = \sum_{t=1}^{N} u(c_t)$ if healthy and $U = h(m) + \sum_{t=1}^{N} u(c_t)$ if sick. The individual chooses their consumption vector c and the level of medical spending m after observing the realization of the health state subject to budget constraints. This choice process is described in more detail below as we describe two types of individuals varying in their liquidity constraints.

Before period 1, the individual enters into an insurance plan $Z = (p, \alpha)$ characterized by a total insurance premium p and proportional insurance level α. That is fraction α of medical spending is covered by insurance, leaving the individual with out-of-pocket costs $(1 - \alpha)m$. Note that the level of insurance α will affect the level of medical spending optimally chosen by the individual– this allows for the possibility of moral hazard. We assume in our baseline analysis that premiums are actuarially fair such that $p = \pi\alpha m$ and reflect the endogenously chosen level of medical spending in the sick state.[9] As in most models incorporating the possibility of moral hazard, the individual

[8] This simplifying assumption eliminates the need to consider more complicated dynamics related to the potential timing of when a sick state might arise during the N periods that would affect consumption decisions for individuals that can borrow and save across consumption periods. For individuals that are fully constrained to live hand-to-mouth and can neither borrow nor save, this assumption can be relaxed without complicating the model. Ericson and Sydnor (2018) show that this assumption– termed perfect foresight– is necessary for an individual with no liquidity constraints to be represented with a standard static expected utility model.

[9] This assumption that premiums reflect the chosen level of medical spending could imply either that they are set with correct expectations of how individuals will respond to their insured incentives or could be thought of as a longer-run equilibrium result where premiums and spending decisions adjust to be mutually consistent.

will choose medical spending taking the level of premium as fixed and not internalizing the impact of medical spending on premiums.

2.2 Types of individuals

We consider two types of individuals $i \in \{PL, HTM\}$. The first, denoted PL, has "perfect liquidity", which means they can allocate their total income toward consumption, premium payments and medical spending across periods, and simply must satisfy a lifetime budget constraint; they can costlessly to borrow from future income, and their savings earn no rate of return.

Given these assumptions the optimal decision for the individual with "perfect liquidity" will be to perfectly smooth consumption across all periods. As such, the ex-ante expected utility for individuals with perfect liquidity is given by:

$$EU_{PL} = \pi \left[Nu\left(\frac{y - (1-\alpha)m(\alpha) - p}{N} \right) + h(m(\alpha)) \right] + (1-\pi) Nu\left(\frac{y-p}{N} \right). \tag{1}$$

This utility function is equivalent to static expected utility model used in prior work on health insurance and moral hazard, including Finkelstein et al. (2019).

The second type of individual, denoted HTM, has liquidity constraints and lives hand-to-mouth – they cannot move resources between periods. They consume all available resources each period.[10] They have to pay any medical spending spread over $K < N$ consumption periods. We assume that premium payments are made smoothly over time, such that they pay a $1/N^{\text{th}}$ of the premium in each period.[11] For these individuals the ex-ante expected utility is given by:

$$EU_{HTM} = \pi \left[(N-K)u\left(\frac{y-p}{N} \right) + Ku\left(\frac{y-p}{N} - \frac{(1-\alpha)m(\alpha)}{K} \right) + h(m(\alpha)) \right] + (1-\pi) Nu\left(\frac{y-p}{N} \right). \tag{2}$$

The key difference from the perfect-liquidity case is that the cost of uninsured medical expenses $(1-\alpha)m(\alpha)$ is born in only K periods instead of being spread across all N periods. $K = 1$ describes the extreme case in which the individual receives a medical treatment in a given consumption period and has to pay for it immediately. $K > 1$ can appear if the medical treatment and associated bills are spread over several periods, or the providers allow for a payment plan of the bills. Setting $K = N$ yields the perfect liquidity case above and has been the implicit assumption in applied work so far.

---

[10] The stark hand-to-mouth case facilitates developing theoretical results. As shown in Ericson and Sydnor (2018), the same insights translate to a more computationally difficult model in which individuals can borrow at high interest rates.

[11] This assumption of smooth premium payments matches the empirical situation that is typical for health insurance contracts in the United States. See Ericson and Sydnor (2018) for a discussion of alternative premium-payment timing.

2.3 Defining Moral Hazard

We distinguish between the *chosen* and *efficient* level of medical spending for a given amount of insurance coverage. The level of medical spending chosen when sick to maximize ex-post utility for an individual of type i with insurance coverage α is $m_i^*(\alpha)$. When an individual has no insurance, their chosen medical spending is denoted $m_i^*(0)$. The individual will select this amount of coverage $m_i^*(\alpha)$ taking the premium cost of insurance as given.

The efficient (i.e., optimal) level of medical spending is denoted by $m_i^E(\alpha)$. This is the level of medical spending that individuals of type i, taking the level of insurance α as given, would commit to *ex-ante* if this commitment was binding and the individuals would internalize the full effect of this medical spending on their premiums. The efficient level of medical spending, for a given level of insurance, is thus the level of m that maximizes expected utility subject to the constraint that $p(m, \alpha) = \pi\alpha m$, that is $m_i^E(\alpha) = \arg\max_m EU_i(m, p(m, \alpha), \alpha)$.

The crucial difference between chosen $m_i^*(\alpha)$ and efficient $m_i^E(\alpha)$ is that $m_i^*(\alpha)$ is chosen holding fixed the level of premiums, while $m_i^E(\alpha)$ is chosen accounting for its effect on the cost of insurance. In the next section we explore the first order conditions that govern the efficient and chosen levels of spending.

Note that the level of efficient spending will depend on the level of insurance. As we will show, if spending were contractible, the amount individuals would contract on would depend on their insurance level. Efficient spending is defined for any insurance level; this definition applies even if the individual does not have the optimal level of insurance (e.g. was assigned to a plan or made a mistake).

Our definition of efficient spending differs from the definition used by Nyman (1999b) and summarized in Nyman et al. (2018). Nyman defines efficient spending as the level that would arise from receiving a lump-sum transfer when sick of the amount the individual would have spent under insurance (see also Chetty, 2008). In Appendix B we present a deeper discussion and derivation of the differences between our measure and the lump-sum measure. Setting the lump sum amount equal to the amount the individual would spend under insurance will not achieve the efficient level of spending, as we define it. The problem is that the total amount being spent when insured is inefficient and hence the lump-sum transfer is too large. Lump-sum contingent transfers could be used in theory to achieve the efficient level of spending, but the amount of the lump sum transfer would need to be chosen to yield $m_i^E(\alpha)$ as the result of the individual's maximization problem. We believe our definition of efficient spending is the correct one, since it is derived from the full internalization of the utility cost of spending given the level of insurance.

Having set up these definitions, we can now define:

1. Observed moral hazard: $MH_i(\alpha) = m_i^*(\alpha) - m_i^*(0)$
2. Efficient moral hazard: $MH_i^E(\alpha) = m_i^E(\alpha) - m_i^*(0)$
3. Inefficient moral hazard: $MH_i^I(\alpha) = m_i^*(\alpha) - m_i^E(\alpha)$

The first empirical object $MH_i(\alpha)$ is the total observed change in medical spending when an individual has insurance level α versus when they are uninsured. The observed moral hazard can be disentangled into an efficient ($MH_i^E(\alpha)$) and inefficient ($MH_i^I(\alpha)$) component. Note that $MH_i(\alpha) = MH_i^E(\alpha) + MH_i^I(\alpha)$. The efficient component of moral hazard is the change in spending that the individual would optimally choose, while the inefficient component of moral hazard is the "excess" spending.

2.4 Defining the Value of Insurance

These concepts allow us to decompose the value of insurance into a component resulting from moral hazard, and a separate component from its effect on risk and financial smoothing. The total value of insurance for an individual of type $i \in \{PL, HTM\}$ at coverage level α is the change in expected utility relative to no insurance:

$$V_i(\alpha) = EU_i(\alpha, m_i^*(\alpha)) - EU_i(0, m_i^*(0)), \tag{3}$$

where each of the expected utility terms are evaluated at the actuarially fair premiums consistent with the chosen level of medical spending ($m_i^*(\alpha)$ and $m_i^*(0)$, respectively).

This total value can be split into the moral hazard value and a "risk and financing value" such that $V_i(\alpha) = V_i^{MH}(\alpha) + V_i^{RF}(\alpha)$, where

$$V_i^{RF}(\alpha) = EU_i(\alpha, m_i^*(0)) - EU_i(0, m_i^*(0)) \text{ and} \tag{4}$$
$$V_i^{MH}(\alpha) = EU_i(\alpha, m_i^*(\alpha)) - EU_i(\alpha, m_i^*(0)). \tag{5}$$

The risk and financing value V_i^{RF} is the change in utility from buying actuarially fair insurance, but holding fixed the medical spending at the uninsured level $m_i^*(0)$. This entails using insurance to reduce risk and finance spending but not to change spending.[12]

The moral hazard value $V_i^{MH}(\alpha)$ is the change in utility from the individual choosing medical spending ex-post with insurance ($m_i^*(\alpha)$) versus having medical spending fixed at the uninsured level $m_i^*(0)$. This value can be further decomposed into the value of efficient moral hazard, V_i^{MHE}, and the value of inefficient moral hazard, V_i^{MHI} as follows:

$$V_i^{MHE}(\alpha) = EU_i(\alpha, m_i^E(\alpha)) - EU_i(\alpha, m_i^*(0)) \text{ and} \tag{6}$$
$$V_i^{MHI}(\alpha) = EU_i(\alpha, m_i^*(\alpha)) - EU_i(\alpha, m_i^E(\alpha)). \tag{7}$$

where such that $V_i^{MH}(\alpha) = V_i^{MHE}(\alpha) + V_i^{MHI}(\alpha)$. While it is typically assumed that $V_i^{MH}(\alpha)$ is negative, our next section will show that it can be positive.

---

[12] As Ericson and Sydnor (2018) discuss and we highlight in the next section, for those with perfect liquidity insurance has only a risk-reducing benefit, but has an additional financing benefit for those with liquidity constraints.

3 Model Results

In what follows, we first begin with an intuitive discussion of how the analysis of moral hazard changes when we consider the liquidity and income-effects of insurance. We develop a set of simple graphical depictions that make it possible to visualize both the moral-hazard and welfare implications of insurance. We then present results of numerical simulations that help to highlight the potential magnitude of the differences in insurance value and specifically the impact of moral hazard. In the final subsection, we present formal theoretical propositions under more general conditions with proofs provided in the appendix. Some readers with less interest in the technical formalities may choose to skip this final subsection without losing intuitions for the core results.

3.1 Graphical Analysis and Intuition

We begin with an intuitive discussion of the key results. We highlight the following key points:

1. The liquidity benefits of insurance lead part of observed moral hazard to be efficient for hand-to-mouth agents. There can be positive moral hazard value from insurance for hand-to-mouth agents even when the sick state is sure to occur.

2. When the sick state is not sure to occur, there is also an income effect of insurance ("de Meza-Nyman income effect"). This leads part of observed moral hazard to be efficient for both perfectly liquid and hand-to-mouth agents.

3.1.1 Illustrating the liquidity effect using the sure-loss case ($\pi = 1$)

When losses are sure to occur ($\pi = 1$), the perfect-liquidity case conforms to the standard economic intuition introduced by Pauly (1968). For those with perfect liquidity, the efficient level of spending is the level that would be chosen when uninsured. To see this, recall the expected utility for the perfect liquidity case from Equation 1:

$$EU_{PL} = \pi \left[Nu\left(\frac{y - (1-\alpha)m(\alpha) - p}{N} \right) + h(m(\alpha)) \right] + (1-\pi) Nu\left(\frac{y-p}{N} \right). \tag{8}$$

First, consider the efficient level of spending. Since premiums are defined by $p = \pi\alpha m$ and $\pi = 1$ here, the above equation reduces simply to:

$$EU_{PL} = Nu\left(\frac{y-m}{N} \right) + h(m). \tag{9}$$

This equation does not depend on the level of insurance α, since any increase in spending must be paid for fully either via uninsured spending or insurance premiums. As such, the first order condition implicitly determining the optimal level of spending when uninsured ($m^*_{PL}(0)$) and the

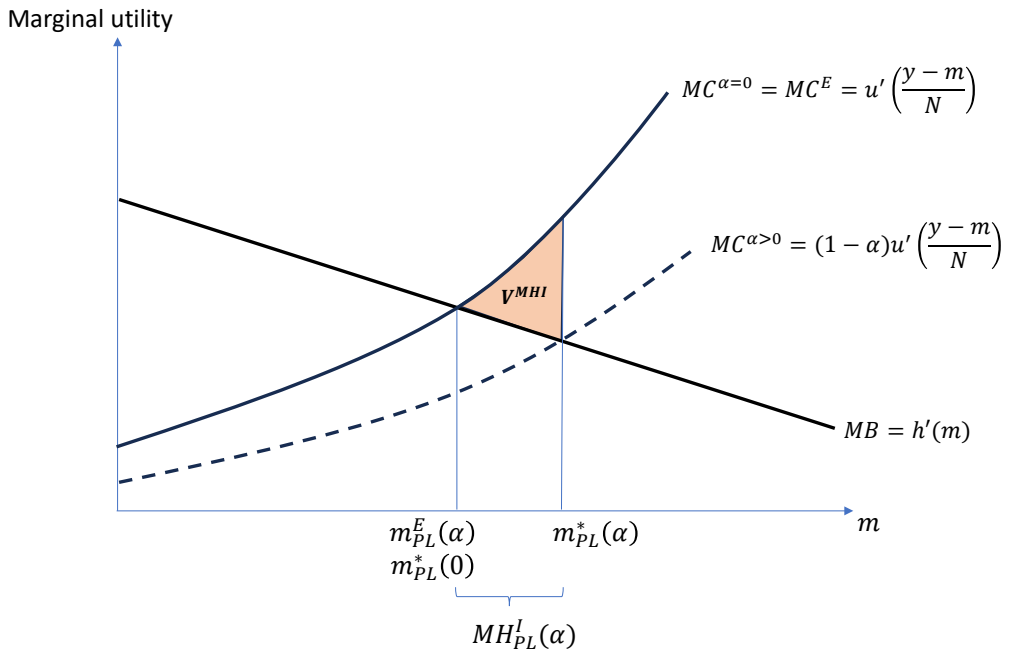
efficient level of spending ($m^E_{PL}(\alpha)$) is the same and given by:

$$FOC^E_{PL} = FOC^0_{PL} : u'\left(\frac{y-m}{N}\right) = h'(m). \tag{10}$$

Insurance, though, reduces the perceived price of spending and leads to inefficient moral hazard. The reason is that the individual takes the premium as given when deciding on the optimal level of insured spending ($m^*_{PL}(\alpha)$), which results in the following first order condition:

$$FOC^*_{PL} : (1-\alpha)u'\left(\frac{y-m}{N}\right) = h'(m). \tag{11}$$

Figure 1 – Standard Analysis of Moral Hazard (Perfect Liquidity)



Note: Schematic illustration. Assumes perfect liquidity and $\pi = 1$.

Figure 1 gives a graphic illustration of this standard case of inefficient moral hazard for an individual with perfect liquidity. The vertical axis is the marginal utility cost and benefit of additional medical spending. The horizontal axis is the level of medical spending. The marginal benefit (MB) of spending is given by $h'(m)$, which is downward sloping given our assumptions on $h(m)$.

We then plot the marginal costs (MC) of medical spending, which is the marginal consumption utility given the level of medical spending (i.e., the left-hand side of the first-order conditions in Equations 10 and 11). These MC lines slope upward due to the concavity of the utility function ($u'' < 0$)). The line $MC^{(\alpha=0)}$ is the marginal utility cost of foregone consumption with medical spending m when uninsured. Next, MC^E is the marginal cost of additional medical spending accounting for the additional premiums induced by that spending. Here it lies directly on top of $MC^{(\alpha=0)}$, as noted above. The uninsured and efficient levels of spending coincide: $m^*(0) = m^E(\alpha)$. The dashed line, $MC^\alpha > 0$, is the marginal utility cost of foregone consumption through out-of-pocket costs given insurance α at medical spending level m. It is shifted down, since insurance means out-of-pocket costs are lower for any given amount of medical spending. As a result, the medical spending $m^*(\alpha)$ chosen when insured is higher, with the additional spending representing inefficient moral hazard ($MH^I_{PL}(\alpha)$). The shaded area represents the negative welfare cost of inefficient moral hazard (V^{MHI}).

The analysis is different for those living hand to mouth. In their case, there is a difference between the marginal consumption-utility cost of medical spending in the uninsured case and the efficient case with insurance. To see this, we recall the expected utility for the hand-to-mouth individuals from Equation (2), which setting $\pi = 1$ and focusing on the full hand-to-mouth case with $K = 1$ we can write as

$$EU_{HTM} = \left[(N-1)u\left(\frac{y-\alpha m}{N}\right) + u\left(\frac{y-\alpha m}{N} - (1-\alpha)m(\alpha)\right) + h(m(\alpha))\right]. \tag{12}$$

The first order condition for medical spending when insured is:

$$FOC^0_{HTM} : u'\left(\frac{y}{N} - m\right) = h'(m). \tag{13}$$

The difference from the perfect-liquidity case in Equation (10) is that those living hand to mouth with $K = 1$ must absorb the entire medical spending in a single period rather than spreading it across N consumption periods.

The first order condition for the efficient spending level for the hand-to-mouth individuals is

$$\begin{aligned} &FOC^E_{HTM} : \\ &\underbrace{(1-\alpha)u'\left(\frac{y-\alpha m}{N} - (1-\alpha)m\right)}_{\substack{\text{Marginal consumption-utility cost of medical spending born through out-of-pocket costs}}} - \underbrace{\alpha\left(\frac{1}{N}u'\left(\frac{y-\alpha m}{N} - (1-\alpha)m\right) + \frac{N-1}{N}u'\left(\frac{y-\alpha m}{N}\right)\right)}_{\substack{\text{Marginal consumption-utility cost of medical spending born through premiums}}} = h'(m). \end{aligned} \tag{14}$$

For hand-to-mouth individuals, the marginal consumption-utility costs of medical spending are lower with insurance, resulting in the efficient level of spending being higher than uninsured spend-

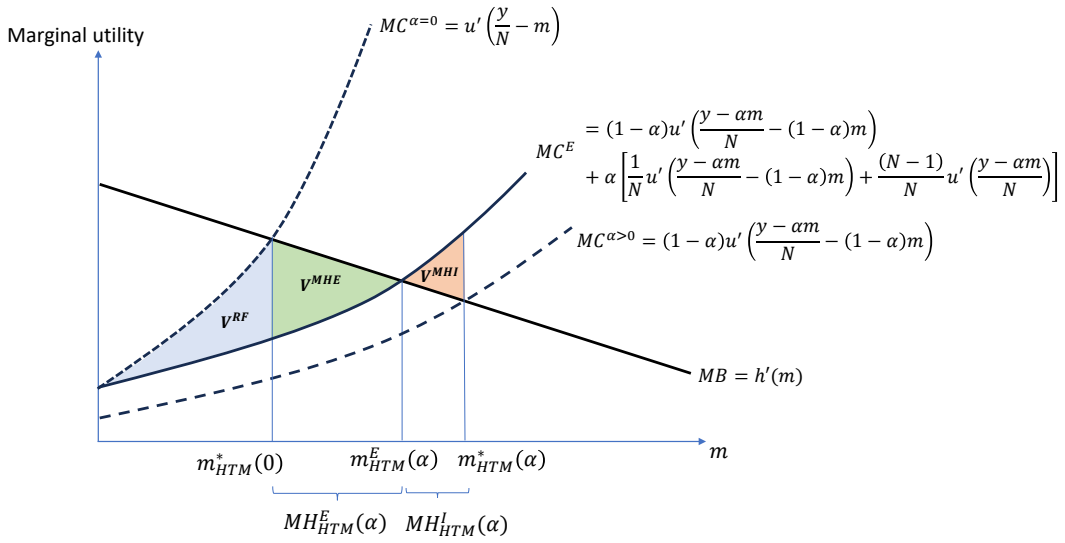
ing. The key reason is that the insured portion of spending, αm, is spread out across all N consumption periods.[13]

At the same time, there will still be some level of inefficient moral hazard in the hand-to-mouth case. When choosing actual medical spending, the individual takes premiums as given. The first order condition for the chosen level of insurance only includes the first term on the left-hand side of Equation 14 and is given by

$$FOC^{*}_{HTM} : (1-\alpha)u'\left(\frac{y-\alpha m}{N} - (1-\alpha)m\right) = h'(m). \tag{15}$$

Figure 2 shows the hand-to-mouth case graphically. The full marginal cost under insurance accounting for its impact on premiums (MC^E) lies in between the uninsured marginal cost and the marginal cost under insurance. This implies that there is both efficient and inefficient moral hazard.

Figure 2 – Analysis of Moral Hazard for Hand-to-Mouth Individuals



Note: Schematic illustration. Assumes hand-to-mouth behavior and $\pi = 1$.

There are now three shaded regions denoting the welfare impacts of insurance and moral hazard specifically. The V^{MHI} area shaded in pink is the same inefficiency of moral hazard as we saw in

[13] As such, the marginal consumption-utility cost in the period experiencing the loss is lower with insurance. In addition, the $(N-1)$ consumption periods where no loss is experienced and only premiums are paid have higher levels of consumption and thus lead to a lower average marginal consumption-utility cost for premiums than for uninsured spending. Mathematically we have that: $\frac{1}{N}u'\left(\frac{y-\alpha m}{N} - (1-\alpha)m\right) + \frac{N-1}{N}u'\left(\frac{y-\alpha m}{N}\right) < u'\left(\frac{y-\alpha m}{N} - (1-\alpha)m\right) < u'\left(\frac{y}{N} - m\right)$.

the perfect liquidity case–resulting from the fact that the level of spending chosen when insured is higher than the efficient level of spending. However, we also have the green-shaded area V^{MHE}, which represents the utility value of the efficient portion of moral hazard – resulting from the fact that the efficient level of spending is higher than the uninsured level. Moreover, we see the shaded dashed area V^{RF} is the risk and financing value. This represents the welfare improvement holding fixed spending at the uninsured level $m^*_{HTM}(0)$ that comes from being able to insure against that spending. It is positive (even though $\pi = 1$) because of the financing benefit of insurance for the liquidity constrained. Each dollar of spending has a lower impact on consumption utility when it is spread out evenly as insurance premiums across N consumption periods.

This illustration shows a visual example where the net welfare value of moral hazard is positive, since the welfare gain from efficient moral hazard is larger than the welfare loss from inefficient moral hazard. This will not always be the case in the hand-to-mouth case, but there will always be an efficient level of moral hazard that helps to offset the inefficient moral hazard.

An intuitive explanation of these results is as follows. For someone who has perfect liquidity, optimization implies that they are indifferent between a dollar more of medical spending and a dollar more of consumption on the margin. Thus, when they become insured and face a lower marginal price, the benefit of medical spending is of lower value than an additional dollar of consumption. The analysis is different for the liquidity constrained. When uninsured, they are indifferent on the margin between a dollar more of medical spending and a dollar of consumption in that period when they are sick. However, forgoing a dollar of premiums spread out over multiple consumption periods is not as painful as forgoing that dollar of consumption in the single sick period. Thus, an additional dollar of medical spending will be more valuable than the cost of financing that with spending with a dollar of premiums. Thus, some of the increased medical spending has higher value to the individual than the full cost to finance it via premiums. Is this fact, that insurance premiums have financing built in, that leads to some of the observed moral hazard being efficient.

3.1.2 Illustrating the income effect using the full-insurance case ($\alpha = 1$)

When the sick state is not sure to happen ($\pi < 1$), there is an additional benefit of insurance that we refer to as the "income effect" following Nyman (1999a).[14] The income effect creates efficient moral hazard for both the hand-to-mouth and perfect-liquidity types. The key idea is that every dollar of insured spending only ends up affecting consumption by its expected costs π via premiums that are shared across those in the sick state (π share) and the non-sick states ($(1-\pi)$ share). We provide formal propositions for the general case in Section 3.3, but here we illustrate the intuition under full insurance.

At full insurance the expected utility coincides for both types. Because premiums are paid smoothly over all N periods and there are no uninsured cost-shocks, the hand-to-mouth individuals

[14] See Appendix B for a discussion of the differences in our definition of efficient moral hazard from the approach used in Nyman (1999a).

in our model become essentially perfectly liquid. That is, we have

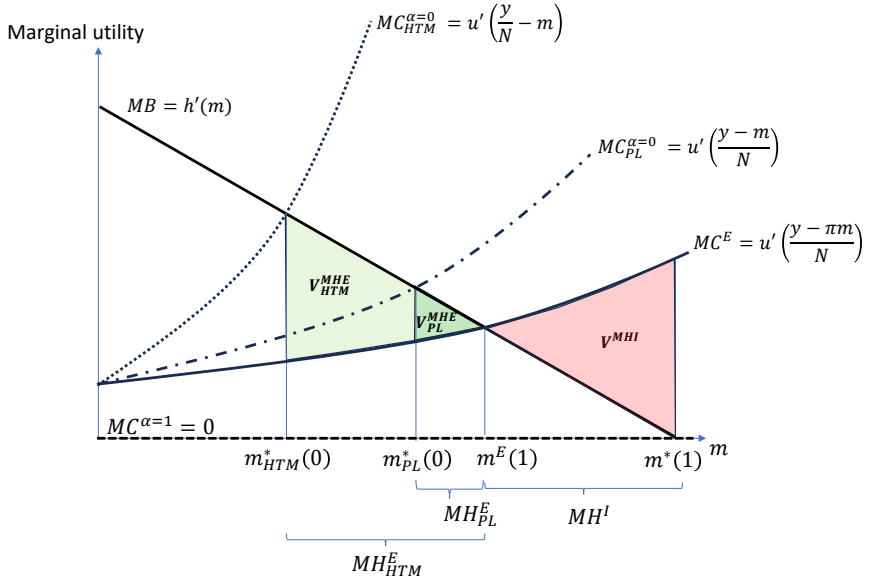
$$EU_{PL}(\alpha = 1) = EU_{HTM}(\alpha = 1) = \pi \left[Nu\left(\frac{y - \pi m(1)}{N}\right) + h(m(1)) \right] + (1 - \pi) Nu\left(\frac{y - \pi m(1)}{N}\right). \quad (16)$$

It is straightforward to show that the efficient level of spending $m^E(1)$, which will be the same for both types, will be govered by the first order condition

$$FOC^E : u'\left(\frac{y - \pi m}{N}\right) = h'(m). \quad (17)$$

Full insurance guarantees that consumption is steady and based on income minus premiums at $\frac{y-\pi m}{N}$ in every period, regardless of whether the loss occurs or not. As such, the efficient spending level is determined by where the marginal consumption utility from paying premiums equals the marginal health benefit. This level of marginal consumption-utility cost is lower than the uninsured levels.

Figure 3 – Analysis of Moral Hazard for Both Types under Full Insurance



Note: Schematic illustration. Assumes full insurance $\alpha = 1$. Note that the subscript PL denotes the perfect-liquidity type and subscript HTM is the hand-to-mouth type. Curves, spending levels, and moral hazard ranges without subscript are the same for both the PL and HTM.

Figure 3 shows the full insurance case graphically. The income effect for the perfect-liquidity case is generated by the difference between the uninsured marginal consumption-utility cost $u'\left(\frac{y-m}{N}\right)$ (Equation (10)) and the efficient marginal cost, $u'\left(\frac{y-\pi m}{N}\right)$. This difference comes from sharing the medical spending costs with uninsured states through premiums. The gap is bigger when the probability of the sick state π is smaller. The efficient level of spending at full insurance $m^E(1)$ is

higher than the uninsured spending, implying that there is efficient moral hazard for the perfect-liquidity case. The positive welfare value of the efficient moral hazard is given by the darker green shaded region V_{PL}^{MHE}. At the same time, because the individual perceives no cost to medical spending under full insurance, they over-consume at the point $m^*(1)$, leading to inefficient moral hazard. In this graphical representation, the net welfare value of moral hazard for the liquidity constrained individual is negative since the losses from inefficient moral hazard are larger than the gains from efficient moral hazard. There can be situations, however, where the efficiency gains from the income effect lead to positive net moral hazard value for those with perfect liquidity. We give conditions for that to happen in the general case in our formal propositions below and provide numerical examples in the final subsection.

The moral hazard benefits of insurance are larger for the hand-to-mouth individuals because they have both the liquidity effect and the income effect of insurance. The liquidity effect can be seen in the figure by comparing the lines for uninsured marginal cost for the hand-to-mouth $u'\left(\frac{y}{N}-m\right)$ to the lower uninsured marginal cost under perfect liquidity. The liquidity effect spreads the medical spending across consumption periods, while the income effect spreads the costs across sick and non-sick states via premiums. The level of efficient moral hazard is greater for the hand-to-mouth case because the uninsured level of spending is lower, and the corresponding size of the welfare gain from efficient moral hazard V_{HTM}^{MHE} (denoted by the larger light green region) is larger. The level of inefficient moral hazard is the same for hand-to-mouth types as it is for those with perfect liquidity because the chosen level of spending is the same under full insurance. As we show in the formal propositions, because the hand-to-mouth types have both the liquidity effect and income effect, the net value of moral hazard is more likely to be positive for the hand-to-mouth types. An implication is that a benevolent social planner would typically set a higher level of insurance for hand-to-mouth types than for perfect-liquidity types.

3.2 Numerical examples

This subsection provides numerical examples to help quantify the plausible size of the welfare effects of moral hazard and to draw out a few other important comparisons related to the results highlighted so far.

We use a set of linear approximations to the marginal consumption-utility costs in the first-order conditions that govern uninsured, efficient, and chosen levels of spending. We use first-order Taylor expansions of the marginal consumption utility terms around the marginal utility at baseline consumption $u'\left(\frac{y}{N}\right)$. We then normalize the entire first order condition by dividing by $u'\left(\frac{y}{N}\right)$. This results in a set of linear marginal (consumption-utility) cost curves that are in monetary units with intuitive functional forms. Appendix C provides the full set of derivations for general levels of insurance and probability of sickness.

The linear approximations to the first-order conditions governing uninsured spending in the perfect-liquidity and hand-to-mouth cases (with $K = 1$) are:

$$1 + r\frac{m_{PL}(0)}{N} \approx H'(m_{PL}(0)) \tag{18}$$

$$1 + rm_{HTM}(0) \approx H'(m_{HTM}(0)). \tag{19}$$

The left-hand side of these equations is the marginal consumption-utility cost of medical spending. The slope of the marginal costs are governed in part by r, which is the Arrow-Pratt measure of absolute risk aversion at baseline consumption ($r = -\frac{u''\left(\frac{y}{N}\right)}{u'\left(\frac{y}{N}\right)}$). We see that without insurance, the marginal cost of medical spending grows with m at a rate of $\frac{r}{N}$ in the perfect-liquidity case and at a rate of r in the hand-to-mouth case. The marginal consumption-utility cost is N times steeper for the hand-to-mouth individuals who have to absorb the entire spending in one period.

Next, examine the first-order condition for efficient spending under full insurance. As we highlighted in the prior subsection, under full insurance, this condition is the same for the perfect-liquidity and hand-to-mouth cases, so we can suppress the type subscript on $m^E(1)$. The first order condition for efficient spending in the linear approximation is:

$$1 + \pi r\frac{m^E(1)}{N} \approx H'(m^E(1)) \tag{20}$$

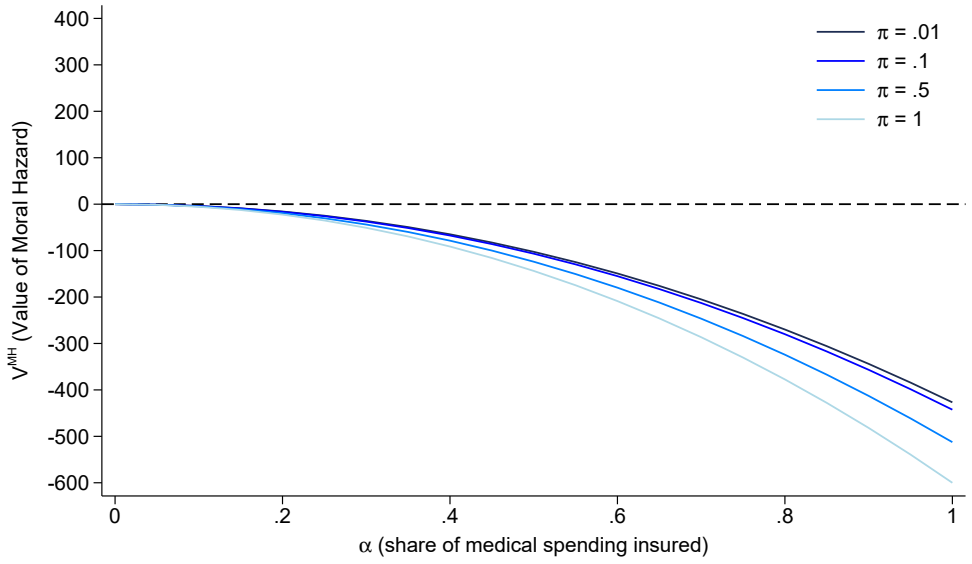
The efficient marginal cost under full insurance rises with m at a rate of $\pi\frac{r}{N}$. The income effect of insurance generates efficient moral hazard for the perfect-liquidity case by lowering the slope of the marginal cost relative to the uninsured case by a factor of π. For the hand-to-mouth case, the efficiency of moral hazard under full insurance arises because the marginal cost under full insurance is reduced by both π and $\frac{1}{N}$.

For our numerical example we pin down these approximate marginal cost curves by assuming $r = 0.0006$. This is the level of absolute risk aversion that corresponds to a Constant Relative Risk Aversion utility function $u = \frac{c^{(1-\rho)}}{(1-\rho)}$ with $\rho = 3$ and evaluated at baseline consumption for annual income $y = \$60,000$ and monthly consumption periods $N = 12$.

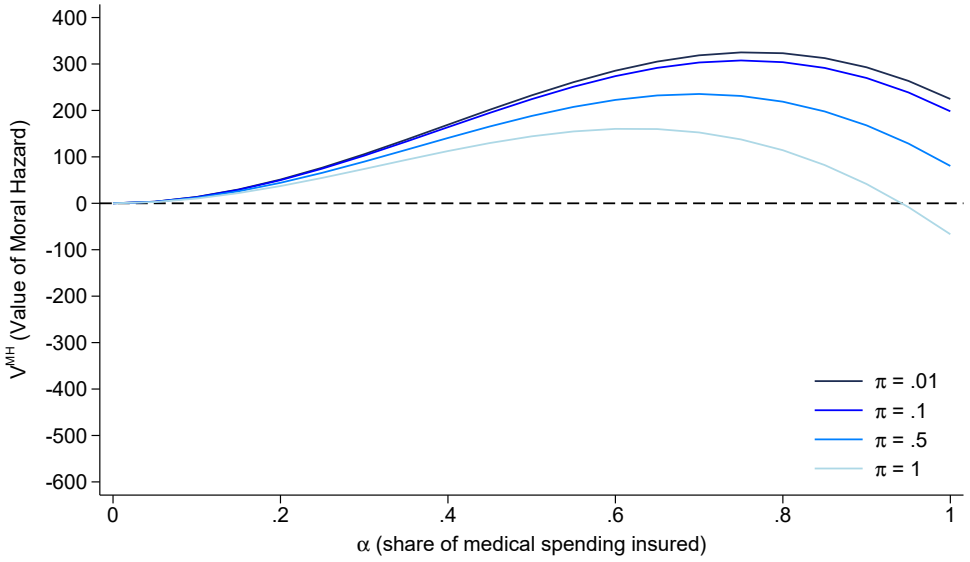
The H' terms on the right-hand side of the first-order conditions are the marginal health benefit of medical spending (h') normalized by dividing by $u'\left(\frac{y}{N}\right)$. We assume that the marginal health benefit is linear. Under this linearity assumption, the H' curve can be pinned down if we know two levels of spending. For our numerical example we assume that uninsured spending for those with perfect liquidity is $m^*_{PL}(0) = \$3,000$ and that the spending by those individuals under full insurance would be $m^*(1) = \$4,000$. With all of the curves pinned down, the welfare values of moral hazard illustrated in the shaded regions of Figures 1 to 3 become triangles that can be easily quantified.

Figure 4 plots the resulting total moral hazard value ($V_i^{MH} = V_i^{MHE} + V_i^{MHI}$) for insurance levels ranging from no ($\alpha = 0$) to full insurance ($\alpha = 1$) and for four different levels of the probability of sick state ranging from a 1 percent chance to a sure sick state. Panel a) plots the results for

Figure 4 – Numerical Examples of Moral Hazard Value



(a) Perfect Liquidity $K = N$



(b) Hand to Mouth $K = 1$

Note: These figures show the net value of moral hazard for different levels of insurance (α) and probability of sick state (π). We use the linear approximation to the marginal utility of consumption outlined in this subsection. The parameters governing the numerical exercise are that a) the number of consumption periods $N = 12$, b) the Arrow-Pratt measure of absolute risk aversion $r_A = 0.0006$, c) the uninsured spending under perfect liquidity is $m^*_{PL}(0) = \$3,000$, and d) the maximum spending observed under full insurance is $m^*(1) = \$4,000$. As described in the text, we assume a linear marginal utility of health function $h'(m) = c - bm$ and the parameters are pinned down by the assumptions about $m^*_{PL}(0)$ and $m^*(1)$. Panel a) shows the perfect liquidity case where $K = N$ and Panel b) shows the full hand-to-mouth case where $K = 1$.

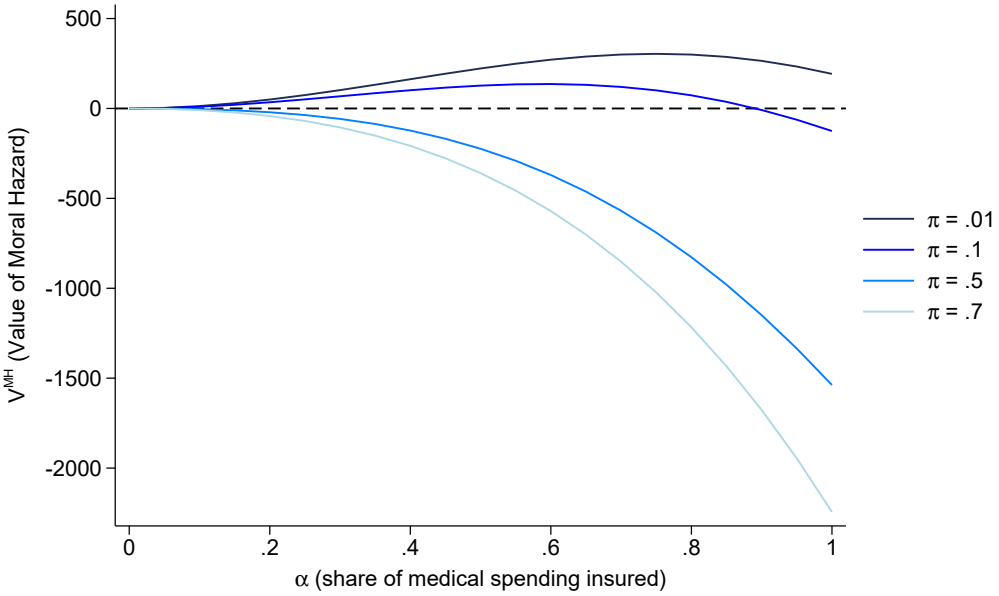
individuals with perfect liquidity. We see that for this numerical example, the total moral hazard value is negative for those with perfect liquidity and worse as the level of insurance gets higher. At higher levels of insurance, moral hazard reduces welfare by around $300 to $600. Due to the income effect there is some level of efficient moral hazard when $\pi < 1$, which can be seen by noting that the total value of moral hazard at every level of insurance is higher for lower values of π. However, the value of efficient moral hazard is low relative to the losses from inefficient moral hazard, leading to negative moral hazard value across the range of values explored here.

For hand-to-mouth individuals, however, the value of moral hazard is often significantly positive. In this example, the value of moral hazard peaks at relatively high levels of insurance, around 80% insurance for probabilities of sickness of 10% or less. Noting that the risk and financing value is rising with the level of insurance, this implies that the optimal level of insurance will be close to full insurance for the hand-to-mouth. The value of moral hazard can be quite significant, reaching over $300 for higher levels of insurance against lower probabilities of sickness. We also note that even when the sick state is sure to arise, the net value of moral hazard is positive up to nearly full insurance (crossing to be negative around 95% insurance). This means that in this example, hand-to-mouth individuals derive positive moral hazard benefits from insurance that helps them finance medical expenditures they are sure to need and that it can be beneficial to finance nearly all such expenditures. This is true even despite the fact that insurance generates some degree of inefficient spending due to distortions in the perceived price of care.

It might be tempting to think about liquidity constraints as simply representing a case of very strong risk aversion, but there is an important distinction between the moral hazard value of insurance under liquidity constraints and under high risk aversion but perfect liquidity. The efficiency gains of moral hazard under insurance arise in the hand-to-mouth case because the marginal consumption-utility cost curve is very steep when uninsured relative to when insured. It is also true that a much higher level of absolute risk aversion can generate a similarly steep marginal cost curve when uninsured for those who have perfect liquidity. Very high levels of absolute risk aversion would arise for those with low annual incomes or when the cost of medical care is high, such that medical spending is traded off against very essential consumption. However, the efficiency benefits without liquidity constraints arise only from the income effect, which depends on a low probability of the sick state. For those with perfect liquidity, as the probability of the sick state approaches 1, the uninsured spending approaches the efficient level of spending. However, for those who are hand-to-mouth, the uninsured spending stays below the efficient level even for $\pi = 1$.

Figure 5 provides an example of this. For this figure we repeat the numerical analysis from Figure 4 for the perfect liquidity case but under the assumption that absolute risk aversion is 12 times the original absolute risk aversion in the prior figure ($r'_A = 12r_A$=0.0072). As such, in this example the uninsured marginal consumption-utility cost for the perfect liquidity individual is now the same as the uninsured marginal consumption-utility cost for the hand-to-mouth in the numerical example in Figure 4 panel b. We use here also the same calibrated marginal utility of health function $H'(m)$ from the prior numerical example, so the uninsured spending in this case

Figure 5 – Numerical Example Perfect Liquidity and High Risk Aversion



Note: This figure shows the net value of moral hazard for different levels of insurance (α) and probability of sick state (π) for the perfect-liquidity case where $K = N$. We use the linear approximation to the marginal utility of consumption outlined in this subsection. The parameters governing the numerical exercise are that a) the number of consumption periods $N = 12$, b) the Arrow-Pratt measure of absolute risk aversion $r_A = 0.0072$, c) the uninsured spending under perfect liquidity is $m^*_{PL}(0) = \$3,000$, and d) the maximum spending observed under full insurance is $m^*(1) = \$4,000$. As described in the text, we assume a linear marginal utility of health function $h'(m) = c - bm$ and the parameters are pinned down by the assumptions about $m^*_{PL}(0)$ and $m^*(1)$.

matches the uninsured spending for the hand-to-mouth individuals from the prior figure. We see now that the net value of moral hazard is positive at low levels of π. In particular, if the likelihood of the sick state were 1%, the highly-risk-averse individual with perfect liquidity derives up to around \$300 in value from moral hazard at moderately high level of insurance and significantly positive value from moral hazard even under full insurance. However, as the likelihood of the sick state rises, the value of moral hazard falls dramatically and becomes substantially negative. Those with high absolute risk aversion benefit from moral hazard when premium costs are low, but as the likelihood of sickness rises and premiums go up, the impact of moral hazard is even worse than it would be if the individual were less risk averse (compare Figure 5 to Figure 4 panel a). In contrast, when the high sensitivity to uninsured spending costs comes from the lack of liquidity, as in Figure 4 panel b, the moral hazard value of insurance comes primarily from its liquidity benefit and is much less sensitive to the chance of the sick state.

3.3 Formal propositions

In this section, we develop a set of formal propositions that describe the behavior of perfect liquidity and hand-to-mouth individuals, and the results that hand-to-mouth individuals have a higher value

of moral hazard and thus a higher level of optimal insurance. All proofs are provided in the appendix.

3.3.1 Perfect-liquidity Individuals

First, Proposition 1 describes individuals with perfect liquidity. Part 1 establishes the result that the efficient level of medical spending increases with the level of insurance. Part 2 confirms the standard intuition that individuals with insurance choose a higher level of medical spending than is efficient. Part 3 establishes the result discussed in Section 3.1.1: when the sick state is sure to happen, the efficient level of spending is the uninsured level of spending. Finally, Part 4 covers a different result, known from Nyman (1999a): the efficient level of medical spending is larger than the level that would be chosen if the individual were uninsured.

Proposition 1. *For individuals with perfect liquidity*

1. $u''' \geq 0$ *is a sufficient condition for* $\frac{\partial m_{PL}^{E}(\alpha)}{\partial \alpha} > 0 \; \forall \; \pi \in]0,1[$,
2. $m_{PL}^{*}(\alpha) > m_{PL}^{E}(\alpha) \; \forall \; \alpha \in]0,1], \pi \in]0,1]$,
3. *if* $\pi = 1$, $m_{PL}^{E}(\alpha) = m_{PL}^{*}(0) \; \forall \; \alpha \in]0,1]$, *and*
4. $u''' \geq 0$ *is a sufficient condition for* $m_{PL}^{E}(\alpha) > m_{PL}^{*}(0) \; \forall \; \alpha \in]0,1]$ *if* $\pi \in]0,1[$.

Next, Proposition 2 examines the value of insurance and its components. Part 1 shows that when losses are certain to appear, individuals with perfect liquidity have a negative total value of insurance: there is no risk and financing value ($V_{PL}^{RF}(\alpha) = 0$) since there is no uncertainty, and the moral hazard value of insurance is negative, as the efficient level of medical spending is the level that would be chosen when uninsured. That is, insurance simply brings a distortion in this case.

If we introduce risk to the decision-situation, two effects appear. Insurance now has a positive risk and financing value (Part 2). Further, there is some level of efficient moral hazard (Part 3). Then, drawing on the result on efficient spending from Proposition 1, Part 4 lets us state the argument of Nyman (1999a) formally: there can be situations in which insurance has a positive moral hazard value. The inequality in equation (21) shows a condition sufficient to yield positive moral hazard value. Note that the condition is more likely to be satisfied if π gets smaller. This reflects the argument in Nyman (1999a) that smaller probabilities of getting sick lead to higher transfers of healthy people to sick people through the insurance system. Finally, Part 5 shows that if the condition from Part 4 holds, there is a range of insurance levels where the total value of insurance is positive.

Proposition 2. *For individuals with perfect liquidity*

1. *For* $\pi = 1$, $V_{PL}^{RF}(\alpha) = 0$, $V_{PL}^{MH}(\alpha) < 0$ *and* $V_{PL}(\alpha) < 0 \; \forall \; \alpha \in]0,1]$.
2. *For* $\pi \in]0,1[$, $V_{PL}^{RF}(\alpha) > 0; \forall \; \alpha \in]0,1]$.

3. *For* $\pi \in]0,1[$, $u''' > 0$ *is a sufficient condition for* $V_{PL}^{MHE}(\alpha) > 0; \forall\ \alpha \in]0,1]$.

4. *For* $\pi \in]0,1[$, *there exists* $\hat{\alpha}$ *such that* $V_{PL}^{MH}(\alpha) > 0; \forall\ \alpha \in]0,\hat{\alpha}]$ *if it holds that*

$$(1-2\pi)u'\left(\frac{y-m_{PL}^{*}(0)}{N}\right) - 2(1-\pi)u'\left(\frac{y}{N}\right) > m_{PL}^{*}(0)(1-\pi)\frac{1}{N}u''\left(\frac{y-m_{PL}^{*}(0)}{N}\right). \quad (21)$$

5. *For* $\pi \in]0,1[$, *there exists* $\hat{\alpha}' > \hat{\alpha}$ *such that* $V_{PL}(\alpha) > 0; \forall\ \alpha \in]0,\hat{\alpha}']$ *if Condition* (21) *holds.*

3.3.2 Hand-to-mouth Individuals

Next, we consider the case of an individual who lives hand-to-mouth. Proposition 3 below begins by analyzing optimal spending levels. Part 1 unsurprisingly shows that insurance leads hand-to-mouth individuals to choose a higher level of medical spending than is efficient. However, Part 2 shows a key difference between this case and perfect liquidity: there is efficient moral hazard even if there is no risk, as the socially optimally level of spending is higher than the level that would be chosen when uninsured even at $\pi = 1$.[15]

Proposition 3. *For hand-to-mouth individuals and with* $\pi \in]0,1]$,

1. $m_{HTM}^{*}(\alpha) > m_{HTM}^{E}(\alpha)\ \forall\ \alpha \in]0,1]$,

2. $u''' \geq 0$ *is a sufficient condition for* $m_{HTM}^{E}(\alpha) > m_{HTM}^{*}(0)\ \forall\ \alpha \in]0,1]$.

The reason for this result is that insurance provides financing to spread medical spending across more consumption periods, allowing spending to be borne in lower marginal utility periods than when uninsured. Both items of the proposition taken together show that for hand-to-mouth individuals, independent of the probability of loss, privately optimal medical spending is higher than the socially efficient medical spending at the same level of insurance, which is in turn higher than the uninsured level of medical spending.

The next proposition considers how the results on efficient spending carry over to the value of insurance and its individual components.

Proposition 4. *For* $\pi \in]0,1]$, *hand-to-mouth individuals have*

1. $V_{HTM}^{RF}(\alpha) > 0\ \forall\ \alpha \in]0,1]$,

2. $u''' \geq 0$ *is a sufficient condition that* $V_{HTM}^{MHE}(\alpha) > 0\ \forall\ \alpha \in]0,1]$.

3. *there exists* $\hat{\alpha} > 0$ *such that for* $\alpha \in]0,\hat{\alpha}]$, $V_{HTM}^{MH}(\alpha) > 0$ *if it holds that*

$$(N-2K\pi)u'\left(\frac{y}{N}-\frac{m_{HTM}^{*}(0)}{K}\right) - 2(N-K\pi)u'\left(\frac{y}{N}\right) > m_{HTM}^{*}(0)\frac{N-K\pi}{K}u''\left(\frac{y}{N}-\frac{m_{HTM}^{*}(0)}{K}\right). \quad (22)$$

---

[15] We assume $u''' \geq 0$, which is sufficient but not necessary. Most commonly used utility functions, including CARA and CRRA, satisfy this condition (Brockett and Golden, 1987).

4. there exists $\hat{\alpha}' > \hat{\alpha}$ *such that for* $\alpha \in]0, \hat{\alpha}']$, $V_{HTM}(\alpha) > 0$ *if Condition* (22) *holds.*

First, Part 1 of Proposition 4 considers the "risk and financing" value that holds fixed medical spending at the uninsured level. It is positive even for $\pi = 1$ when there is no uncertainty and thus no standard risk-reduction value from insurance. It becomes positive for the same reason efficient spending with insurance is higher than $m^*_{HTM}(0)$ even if $\pi = 1$: insurance enables the financing of medical payments across all periods, allowing expenses to be borne in periods with lower marginal utility of consumption.

We now consider the moral hazard induced by the insurance coverage. We show in Part 2 of Proposition 4 that individuals with liquidity constraints have a positive value of efficient moral hazard. Part 3 shows a sufficient condition for the entire moral hazard value of insurance to be positive (note this result does not rely on $\pi < 1$ and it is independent of that shown by Nyman, 1999a). Finally, when the value of moral hazard is positive, it also the case there is a range of insurance levels where the total value of insurance is positive (Part 4).

3.3.3 Comparing the Types of Individuals

In what way does living hand-to-mouth change the value of insurance for individuals? Propositions 1 to 4 answer this question for the case of certain losses. When there is no uncertainty, under perfect liquidity, insurance only decreases the welfare of the individual. For hand-to-mouth individuals, on the other hand, the value of insurance can indeed be positive. Intuitively, hand-to-mouth individuals benefit from the financing provided by insurance coverage while this benefit does not exist for perfect liquidity.

When introducing risk, the comparison gets more complicated. There is now the potential of efficient moral hazard for individuals with perfect liquidity due to the effect described by Nyman (1999a): insurance serves as a redistribution system from the healthy to the sick, making consumption higher in the sick state. This state has high marginal utility of consumption for uninsured individuals because medical spending has lowered consumption. Thus, the redistribution increases expected utility. This effect gets stronger with lower loss probabilities and higher medical spending. Because medical spending is higher for individuals with perfect liquidity, this effect is larger for them than it is for hand-to-mouth individuals.

We can nevertheless make certain comparisons. Proposition 5 quickly results from the different ways in which individuals can finance their medical spending. At any level of positive cost-sharing (including being fully uninsured), hand-to-mouth individuals will choose less medical spending than perfectly liquid individuals:

Proposition 5. *For* $K < N$, $m^*_{HTM}(\alpha) < m^*_{PL}(\alpha)$ *for all* $\alpha \in [0, 1[$.

At full insurance, the experience of hand-to-mouth and perfectly liquid individuals is the same– smooth premium payments and no out-of-pocket medical costs. Thus, the chosen level of medical spending will be the same for the two types ($m^*_{HTM}(1) = m^*_{PL}(1)$), as well the efficient level of medical spending ($m^E_{HTM}(1) = m^E_{PL}(1)$). Then, Corollary 1 follows from the definitions of V^{MH}

and V^{MHE}, and says that for the hand-to-mouth individual, the value of both moral hazard and efficient moral hazard, will be higher than for the perfectly liquid individual.

Corollary 1. *For $K < N$ and $\alpha = 1$, $V^{MHE}_{HTM} > V^{MHE}_{PL}$ and $V^{MH}_{HTM} > V^{MH}_{PL}$.*

While Corollary 1 shows that, at full insurance, the value of moral hazard is higher for hand-to-mouth individuals, that value may still be negative. The next proposition considers the range of insurance values for which the value of moral hazard is positive. After making a simplifying assumption on the health benefits function, Proposition 6 shows that hand-to-mouth individuals have a larger range of insurance values for which moral hazard value is positive, than do perfect liquid individuals.

Proposition 6. *Assume that $K < N$, $u''' \geq 0$, and that benefits from medical spending are approximately linear in the relevant ranges of m_i. If for both types there exists an $\hat{\alpha}_i \in]0,1[$ such that $V^{MH}_i(\alpha) > 0$ if and only if $\alpha \in]0, \hat{\alpha}]$, then $\hat{\alpha}_{HTM} > \hat{\alpha}_{PL}$.*

Finally, we consider the optimal level of insurance α^*_i that an individual would choose for themselves, given that they faced actuarially fair premiums. (This is also the level of insurance that a social planner, with the goal to maximize individual welfare, would choose.) We define this optimal level as $\alpha^*_i = \arg\max_{\alpha \in [0,1]} EU_i(\alpha)$ subject to the constraint that premiums $p = \alpha \pi m^*_i(\alpha)$.

Proposition 7 shows that the optimal level of insurance coverage is higher for hand-to-mouth individuals than for those with perfect liquidity. (For an extended discussion see Appendix D.)

Proposition 7. *For $K < N$, if benefits from medical spending are approximately linear in the relevant ranges of m_i and the optimal insurance coverage is characterized by an interior solution, then $\alpha^*_{HTM} > \alpha^*_{PL}$.*

4 Revising Estimates of the Value of Medicaid

4.1 Model

To assess how liquidity constraints impact our interpretation of moral hazard, we adapt the framework developed by Finkelstein et al. (2019) (hereafter, FHL) for evaluating the welfare consequence of the Medicaid expansion using data from the Oregon Health Insurance Experiment. When we adapt their framework to our model with monthly consumption periods, we reproduce their results when $K = N$ and individuals have perfect liquidity. However, derive new adjustments to their key equations when $K < N$ and individuals have liquidity constraints (i.e. live hand-to-mouth). We use those adjustments and data from the Oregon Health Insurance Experiment to re-estimate the value of a year of Medicaid to recipients, focusing specifically on how assumptions about liquidity affect individuals' implied valuation of empirical moral hazard.

We follow the "consumption-based optimization approach" laid out by FHL (they lay out a variety of other approaches as well). Our baseline model and FHL's consumption-based optimization approach share common assumptions: people select a continuous level of health spending given a

realized health shock by optimally trading off medical spending and consumption, which is separable between health and consumption. FHL discuss limitations of approach: e.g. individuals may make mistakes in allocating health spending, or health spending may be lumpy such that the first order conditions do not hold. These limitations apply to our analysis as well. Neither our results below nor the original FHL analysis using this approach are perfect measures of the value of Medicaid to recipients. Our goal, however, is to examine how the estimates would change if Medicaid recipients live hand-to-mouth rather than having perfect liquidity.

We use the same notation as introduced in Section 2 and only adjust it where necessary to fit the new context. In the FHL model, individuals have utility over consumption c and health h, where health is a function of medical spending m: $U(c,h) = u(c) + h(m)$. Health status is indexed by θ.[16] Finally, the level of insurance is $\alpha \in \{0,1\}$, where $\alpha = 1$ is insured and $\alpha = 0$ is uninsured.

We must introduce one additional concept to relate our analysis to FHL. While up until now, the out-of-pocket price of medical care has simply been $1-\alpha$, FHL account for the potential of charity care. Thus, out-of-pocket spending on medical care is $x(\alpha;\theta) = \rho(\alpha)m(\alpha;\theta)$, where $\rho(\alpha)$ is the effective price to the individual. FHL find that charity care covers 79% of expenses, so implement $\rho(0) = 0.21$; at full insurance, $\rho(1) = 0$. In the FHL model (and thus ours), moving from being uninsured to having Medicaid is, from the individual's perspective, akin to moving from 79% insurance to 100% insurance.[17]

The key object in the FHL setup is $\gamma(\alpha)$, which is a measure of Medicaid recipients' willingness to pay for insurance and can be thought of as the premium that if paid would lead to the same utility level as being uninsured. Let $V(\alpha)$ be the indirect utility of insurance level α:

$$V(\alpha) = \max_{m(\alpha;\theta)} \mathbb{E}_\theta \left[(N-K)u\left(\frac{y-\gamma(\alpha)}{N}\right) + (K)u\left(\frac{y-\gamma(\alpha)}{N} - \frac{x(\alpha;\theta)}{K}\right) + h(m(\alpha;\theta)) \right]$$

Then, $\gamma(\alpha)$ is defined so that $V(\alpha) = V(0)$. Note that the equation for $V(\alpha)$ is almost identical to that in FHL, except that we let the payments (γ) be spread across N periods and let the annual medical spending be spread across K periods. When $K = N$ (perfect liquidity), our version reduces to the FHL model. When $K = 1$, we have the hand-to-mouth case.

The value of K in this analysis can be thought of as a joint assumption about the arrival of medical bills over time and the individual's underlying liquidity position. When $K = N$, we are assuming equivalently that either individuals can fully smooth their consumption through costless

---

[16] Consistent with our modeling framework and the implicit assumptions of static models like those used in FHL, we assume that the realization of the health state occurs at the start of the insured period so that the individual with perfect liquidity can fully smooth consumption across periods in the insured year.

[17] In practice, marginal prices may vary depending on the severity of the health shock, the magnitude of expenditure, and the individual's liquidity constraints. If hand-to-mouth individuals faced lower prices when uninsured, they would benefit less from becoming insured . However, perfectly liquid individuals must then have faced a higher uninsured price to maintain the same average. We explore the sensitivity of our estimates to allowing for differences between hand-to-mouth and perfect liquidity individuals, and find that they are relatively insensitive. Appendix E provides details).

borrowing and saving during the insured year or that bills for medical spending are spread out evenly throughout the year on a full payment plan. When $K = 1$, we are assuming that the individual must pay for all medical spending in a single consumption period and is unable to borrow from or save in other consumption periods to help absorb the shock. Neither the implicit assumption of perfect liquidity in FHL ($K = N$) nor the full hand-to-mouth case $K = 1$ is perfectly accurate. In reality, medical spending shocks can arrive in a variety of temporal patterns and sizes throughout the year. In addition, even those with strong liquidity constraints may be offered payment plans and grace periods that provide some degree of financing for medical bills; we discuss how we account for that in the section below. The results for $K = N$ using the original FHL approach and for $K = 1$ for hand-to-mouth provide two different views of the potential value of Medicaid to recipients.

To determine willingness to pay for insurance, $\gamma(1)$, FHL approximate the integral of the marginal value of insurance at each point. We follow the derivation in FHL, adapted to the possibility of spreading premiums and medical expenses across different numbers of periods. Then willingness to pay, γ, changes with insurance in the following way:

$$\begin{aligned}\frac{d\gamma(\alpha)}{d\alpha} = &\frac{\mathbb{E}[u'_L]}{\mathbb{E}\left[\frac{N-K}{N}u'_{NL} + \frac{K}{N}u'_L\right]}(\rho(0) - \rho(1))\mathbb{E}[m(\alpha;\theta)] \\ &+ \text{Cov}\left[\frac{u'_L}{\mathbb{E}\left[\frac{N-K}{N}u'_{NL} + \frac{K}{N}u'_L\right]}, (\rho(0) - \rho(1))m(\alpha;\theta)\right], \end{aligned} \tag{23}$$

where u'_{NL} is the marginal utility of consumption in the consumption periods where no out-of-pocket costs are incurred, so $u'_{NL} = u'\left(\frac{y-\gamma(\alpha)}{N}\right)$. Similarly, u'_L is the marginal utility of consumption in the periods where out-of-pocket costs are incurred, so $u'_L = u'\left(\frac{y-\gamma(\alpha)}{N} - \frac{x(\alpha;\theta)}{K}\right)$. Note that when $K = N$, Equation (23) reduces to FHL's Equation (13) with income and spending values translated to the consumption period (i.e., monthly) level.

When accounting for hand-to-mouth behavior with $K < N$, we see in Equation (23) that the first term is multiplied by a ratio of marginal utilities. The numerator, $\mathbb{E}[u'_L]$, is the expected marginal utility in periods when out-of-pocket medical costs are paid– that is, the benefit of lowering out-of-pocket spending via insurance. The denominator, $\mathbb{E}\left[\frac{N-K}{N}u'_{NL} + \frac{K}{N}u'_L\right]$, is the expected average marginal utility across all periods – that is, the utility cost of that insurance born across all consumption periods via increased premiums. This multiplier can be quite large, because the expected marginal utility in the one period when people must pay for losses can be much larger than the average marginal utility. The second term in Equation (23) will also likely be larger than its perfect liquidity counterpart, as the first term in the covariance, the ratio of marginal utility in the periods suffering a loss to expected marginal utility, will be larger.

Finally, to get the value of full insurance, we follow FHL and integrate this equation over α from 0 to 1. Since intermediate values of α are not observed, they approximate that value by assuming $\gamma(1) \approx \frac{1}{2}\left(\frac{d\gamma(0)}{d\alpha} + \frac{d\gamma(1)}{d\alpha}\right)$. We do the same here. Using this approximation and reflecting the full coverage from Medicaid by setting $\rho(1) = 0$, we can express the total willingness to pay for

Medicaid as

$$\gamma(1) = \underbrace{\rho(0)\mathbb{E}[m(0;\theta)]}_{\gamma^T} + \underbrace{\frac{\mathbb{E}[u'_L] - E[u'_{NL}]}{\mathbb{E}\left[\frac{N-K}{N}u'_{NL} + \frac{K}{N}u'_L\right]}\frac{N-K}{N}\rho(0)\mathbb{E}[m(0;\theta)]}_{\gamma^F}$$
$$+ \underbrace{\frac{1}{2}\text{Cov}\left[\frac{u'_L}{\mathbb{E}\left[\frac{N-K}{N}u'_{NL} + \frac{K}{N}u'_L\right]}, \rho(0)m(0;\theta)\right]}_{\gamma^R}$$
$$+ \underbrace{\frac{\mathbb{E}[u'_L]}{\mathbb{E}\left[\frac{N-K}{N}u'_{NL} + \frac{K}{N}u'_L\right]}\frac{1}{2}\rho(0)\mathbb{E}[m(1;\theta) - m(0;\theta)]}_{\gamma^{MH}}. \tag{24}$$

Here, the first term, γ^T, evaluates the transfer of resources due to insurance. The second term, γ^F, constitutes the financing value of insurance and is unique to hand-to-mouth individuals; it is zero when $K = N$. A share of $\frac{N-K}{N}$ of the expected loss is distributed from the loss consumption periods to the no loss consumption periods. γ^F results from multiplying this share by the difference in marginal utility between these two consumption periods. The third term, γ^R, reflects the risk reduction function of insurance.[18] The last term, γ^{MH}, evaluates the increase in medical consumption due to insurance and thus corresponds to the moral hazard value of insurance.

4.2 Large Bills and Payment Plans

Medicaid bills larger than the resources available to a person cannot be paid; individuals must have some baseline consumption. To account for this, FHL guarantee each individual a consumption floor of minimal consumption, regardless of the medical bills they face. We adapt their approach to an environment with liquidity constraints.

For the perfect liquidity case, we follow FHL and impose an annual consumption floor of \$1,977 (relative to starting annual income per person of \$9,505). As such, the largest medical spending loss for the uninsured is capped at about \$7,500 out of pocket.

However, things are more complicated with liquidity constraints. With $N = 12$ periods, an individual at the annual consumption floor would have monthly consumption \$164.75. To maintain this consumption floor, the maximum medical spending an individual can bear in a given consumption period is \$625. The distribution of annual out-of-pocket medical expenses from FHL, though, has a fair number of losses much higher than \$625 (all the way to \$7,500).

---

[18] Analogous to here, the risk and financing value of insurance, V_i^{RF} as introduced in Section 2.4, can also be decomposed into a risk value and a financing value. To do so, we can introduce a hypothetical insurance contract which concentrates all premium payments into the K loss consumption periods. This would lead to expected utility $EU_i^H(\alpha, m) = \pi\left[(N-K)u\left(\frac{y}{N}\right) + Ku\left(\frac{y}{N} - \frac{p+(1-\alpha)m}{K}\right) + h(m(0))\right] + (1-\pi)\left[(N-K)u\left(\frac{y}{N}\right) + Ku\left(\frac{y}{N} - \frac{p}{K}\right)\right]$. We can then define $V_i^R(\alpha) = EU_i^H(\alpha, m_i^*(0)) - EU_i(0, m_i^*(0))$ and $V_i^F(\alpha) = EU_i(\alpha, m_i^*(0)) - EU_i^H(\alpha, m_i^*(0))$. This differentiation has the sensible properties $V_i^R + V_i^F = V_i^{RF}$, $V_i^R \geq 0 \ \forall \ i$, $V_{HTM}^F > 0$, $V_{PL}^F = 0$, and $V_i^R(\alpha) = 0 \ \forall \ \alpha, i$ if $\pi = 1$.

Thus, we model individuals with liquidity constraints as having access to a "payment plan". We assume that in cases where the individual faces a medical bill over \$625, the losses "spill over" to the next period. So for example, take the $K = 1$ case. If annual spending is \$1,500, then we would have two periods of medical spending of \$625, one period with medical spending of \$250, and 9 periods of no medical spending. Nevertheless, we always assume that the relevant marginal utility then is the marginal utility at that per-period consumption floor.[19]

Thus, even hand-to-mouth individual have access to liquidity in this model; we term the case $K = 1$ as "minimal liquidity." The fraction of individuals who take advantage of these payment plans is non-trivial: 25.9% of individuals in our data will face a medical bill larger than the 1-month consumption ceiling and 8.8% face a medical bill larger than the 3-month consumption ceiling. This implicitly allows for a degree of financing or payment plans for larger hospital bills in cases where the medical spending pushes the individual to the consumption floor. This is a rough approximation to hospital financing plans provided to the uninsured with low incomes. In practice, hospital financing may be more generous in providing a longer smoothing period, but may be less generous in that they still may demand large upfront payments from the uninsured before elective care is given.

4.3 Data

In order to estimate the values in Equation (23), it is necessary to have data on medical spending under Medicaid and for the same population when uninsured. FHL use data from the Oregon Health Insurance Experiment on medical spending and the randomization within that experiment to estimate the difference in spending due to insurance.

For our multiplier on the transfer term, it is necessary to observe the distribution of out-of-pocket spending for the Medicaid population when uninsured. In addition, to estimate both the original "pure insurance" term in FHL and our modification of it, it is necessary to have data on the joint distribution of consumption and out-of-pocket spending for the uninsured.

FHL use two approaches for this and we follow their simpler approach, which they call the "consumption proxy approach". In this approach, they start with an average value of consumption for the low-income uninsured from the Consumer Expenditure Survey (CEX), that they estimate at \$9,214 for the time period and sample they use. They then add to that the average per-capita out-of-pocket spending for the uninsured in the Oregon Health Insurance Experiment to get a total baseline average income level of $y = \$9{,}505$. They then use the distribution of reported out-of-pocket spending for the uninsured in the Oregon Health Insurance Experiment and assume that the only variation in consumption comes from absorbing these out-of-pocket costs when uninsured.

[19] We implement this last assumption to avoid issues with "loss periods" having different consumption levels and hence different marginal utilities. This is in line with a similar assumption in FHL by which they assume that medical losses in excess of \$7,500 lead to the same marginal utility as \$7,500 even though one could assume that such losses are never paid by the individual and thus have no marginal effect on consumption utility.

With Medicaid offering full insurance, the assumption in FHL is that the fully insured will all have fixed annual consumption at the baseline $9,505.

The resulting distribution of consumption per year for the uninsured is plotted as Figure 1 in FHL. We used the replication packet available through the *Journal of Political Economy* to extract the underlying distribution used to make that plot. We use that information and the assumption about baseline income from FHL described above to estimate the values for γ and its decomposition into subcomponents. We follow FHL in assuming that the consumption utility function takes the constant relative risk aversion (CRRA) form with coefficient of relative risk aversion of 3. We further estimate $\mathbb{E}[m(1;\theta) - m(0;\theta)]$ using a counterfactual TSLS approach, as done by FHL. In it, the treatment in the Oregon Health Insurance Experiment serves as the instrument for Medicaid coverage.

4.4 Estimates

Our main results on how liquidity constraints affect the value of Medicaid are given in Table 1. The first column shows the results for the perfect liquidity $K = N$ case using our data and estimation. The next column shows a case of partial liquidity with $K = 3$, and the last with minimal liquidity and $K = 1$.

Table 1 – Impact of Liquidity Constraints on Value of Medicaid

	Perfect Liquidity $K = 12$	Partial Liquidity $K = 3$	Minimal Liquidity $K = 1$
γ, Estimated WTP	2447	4170	5307
γ^T, WTP for transferred resources	605	605	605
γ^F, WTP for financing motive	0	683	1838
γ^R, WTP for risk motive	1750	2678	2467
γ^{MH}, WTP for moral hazard motive	92	203	397
Moral hazard cost at uninsured prices	185	185	185
Ratio of γ_{MH} to moral hazard cost	50%	110%	215%

Note: Source: Authors' calculations from Finkelstein et al. (2019) data. All values are in $US and correspond to the values given in Equation (24). As in FHL, we set $\rho(0) = 0.21$, and assume that consumption utility has a CRRA utility function with a coefficient of relative risk aversion equal 3. N is always set to 12. Uninsured out-of-pocket spending $\rho(0)m(0;\theta)$ is distributed according to the calibrated distribution for out-of-pocket expenses by FHL. $\mathbb{E}[m(1;\theta) - m(0;\theta)]$ is calculated based on a counterfactual TSLS estimation. Both Partial Liquidity and Minimal Liquidity models allow individuals to access payment plans for large bills as discussed in Section 4.2.

Under perfect liquidity, we estimate an overall willingness-to-pay of $2,447. This matches closely the number FHL report under similar assumptions.[20] The careful reader will note, though, that our number is higher than FHL's headline number of $1,421. While their headline number is generated

[20] See FHL's Appendix Table 4. Small differences persist, because our Matlab implementation uses a slightly different numerical procedure than the original paper.

under the assumption of a model of household risk and income sharing, all of our results assume the individual in question bears both the full benefits and the full costs of Medicaid. This allows us to avoid difficult decisions about how to account for children in household risk sharing.

We next examine the decomposition of the estimated willingness-to-pay in the perfect liquidity case. The largest portion ($1,750) is attributable to the reduction in risk faced by the individual, followed by $605 in transferred resources (the expected change in out-pocket medical spending). There is no financing motive, as the individual has perfect liquidity.

Finally, willingness-to-pay for the induced utilization is 50% on the dollar. Access to medicaid led to an average increase in medical spending of $879 compared to being uninsured. When uninsured, the out-of-pocket price for this spending would have been $0.21 \times 879 = \$185$ due to charity care. The individual's value for this additional spending that would have been $185 out of pocket is only 92.3. The final row in the table thus shows the willingness-to-pay for the additional spending divided by the cost of the additional spending to the individual when uninsured, which here is 50%– the standard case of wasteful moral hazard.

Once we leave the assumption of perfect liquidity and consider $K < N$, the estimated willingness to pay for Medicaid increases strongly. We discuss the results for the minimal liquidity ($K = 1$) case; the partial liquidity ($K = 3$) is simply intermediate. The total willingness to pay for individuals with minimal liquidity is more than twice as large as the original estimate. As explained above, this is because Medicaid provides funds in consumption periods when marginal utility is very high and consumers value this transfer accordingly.

Examining the decomposition, the transferred resources component is the same. What is different with minimal liquidity, is a substantial willingness to pay from the financing motive. The willingness to pay from the risk motive is also increased under minimal liquidity.

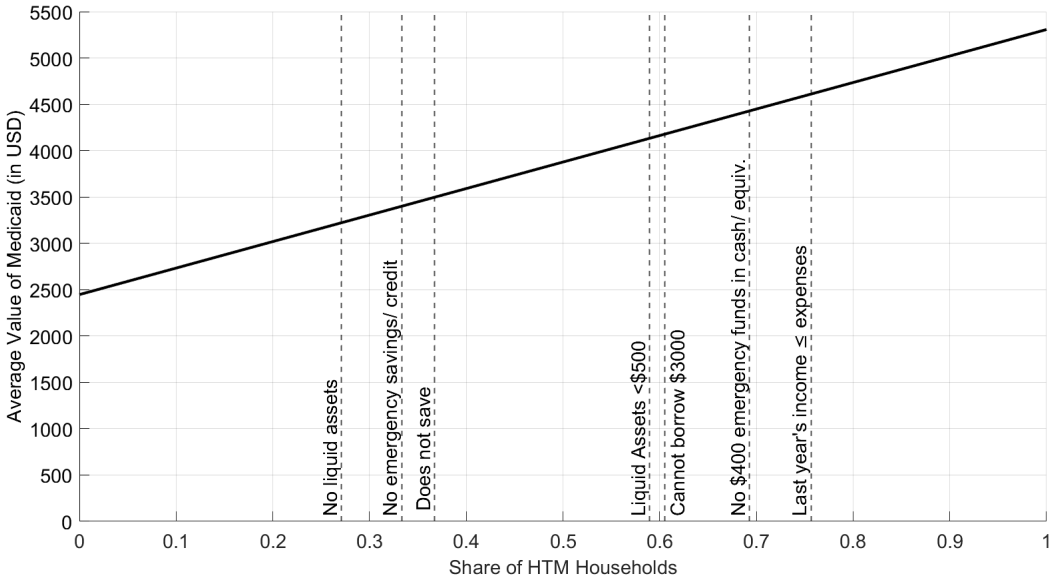
Turning to moral hazard, we see the willingness to pay for the moral hazard motive is more than four times higher under minimal liquidity than in the perfect liquidity case. This is a reflection of our earlier argument: with hand-to-mouth individuals, moral hazard does not have to be an inefficient adjustment, but can rather reflect the possibility of financing health care costs across more consumption periods than the one in which the costs appear.

To interpret the magnitude of the moral hazard results, the individual with minimal liquidity is willing to pay $397 for the additional medical spending that would have cost them $185 out-of-pocket when uninsured; they value this spending at more than 200% of its cost, but did not spend it when uninsured because they could not pay for it in a smooth way.

These results show that the value of Medicaid is substantially higher if the recipients of Medicaid are liquidity constrained. Evidence on liquidity-sensitivity of medical utilization (e.g. Gross et al., 2022) suggests that these constraints do in fact matter.

While we lack data on the full distribution of liquidity constraints and borrowing costs that this population faces, we can provide a range of estimates. Figure 6 illustrates how the average value of Medicaid changes depending on the assumed fraction of Medicaid recipients that is hand-to-mouth versus perfectly liquid. Using data from the Survey of Consumer Finances and Survey of Household

Figure 6 – Value of Medicaid by Share of Hand-to-Mouth Households



Note: The figure displays the value of γ, the estimated willingness to pay for Medicaid, as it is reported in the top row of Table 1 for possible shares of hand-to-mouth ($K = 1$) households in the population. All other households are assumed to have perfect liquidity ($K = 12$). The vertical lines show the values of potential indicators for this share from the Survey of Consumer Finances and the Survey of Household Economics and Decision-making as they are defined in Appendix F.

Economics and Decision-making, we estimate what fraction of the population below 138% of the federal poverty level (a population similar to those eligible for Medicaid) satisfies various conditions that proxy for being liquidity constrained. The vertical lines indicate the fraction of individuals meeting criteria such as having zero liquid assets (27%), having less than $500 in liquid assets (59%), or lacking $400 in emergency savings in cash or cash equivalents (69%). Appendix F provides more detail on these measures. Note as another point of comparison that Aguiar et al. (2024) classify about 40% of the PSID (all income levels) as hand-to-mouth, and that Appendix F shows that low-income individuals are much more likely to display indicators of hand-to-mouth status.

The valuation of Medicaid depends heavily on the fraction of individuals subject to liquidity constraints. For instance, if individuals with zero liquid assets are assumed to be hand-to-mouth (and all others have perfect liquidity), Medicaid's average value is about $3300, whereas assuming those without $400 in emergency funds behave as hand-to-mouth yields an average valuation of roughly $4500. The variation, and the result that the interpretation of moral hazard's welfare impact changes under the hand-to-mouth assumption suggests there would be substantial policy value to additional data on the recipients' of different government insurance programs liquidity constraints and the underlying patterns of medical spending and consumption so that policy recommendations can reflect the value of insurance under liquidity constraints.

Finally, note that our results will depend on the distribution of out-of-pocket medical expenditure across time for the low-income uninsured. We do not know this distribution, but it is likely to be more concentrated in time than the spending of the insured, as uninsured people will screen out lower value care and spend more on urgent health needs. With $K = 1$, our payment plan approach implies that the average fraction of expenses in the highest expenditure month is about 0.77. We can compare this to the distribution of healthcare spending across time in a wealthier sample of people who have employer-sponsored insurance. We use the 2018 Truven Marketscan sample of 24-64 year olds continuously enrolled in insurance (as described in Ericson and Sydnor Ericson and Sydnor, 2018). In this sample, the average fraction of total expenses in the highest expenditure month is 0.54; for smaller expenses, that is higher, about 0.67 for expenses under $1000. Actual OOP expenses are likely more concentrated overtime (e.g. once an individual reaches the maximum out-of-pocket amount, there will be no more cost-sharing), but that will depend on the insurance plan an individual is in. We do not attempt to model those insurance plans because what matters for our sample is the structure of charity care provided for the uninsured, about which there is little data. We thus think our payment plan approach is a reasonable approximation to modeling the concentration of expenses across time for the uninsured.

5 Conclusion

This paper reexamines the interpretation of observed moral hazard from health insurance in the presence of liquidity constraints. We introduce a framework that decomposes moral hazard into efficient and inefficient components for a given level of insurance coverage. In contrast to the traditional view of moral hazard as being wasteful and valued at less than cost, we highlight that because insurance effectively offers a way to smoothly finance lumpy expenditures, the socially optimal amount of spending will increase when an individual with liquidity constraints gets access to insurance.

Accounting for liquidity constraints is important when evaluating the value of expanding insurance programs, particularly for low-income populations. Our reevaluation of the value of Medicaid expansion shows that liquidity constrained recipients would receive a much higher benefit than previously estimated. This would affect the normative evaluation of the expansion, underscoring the importance of accurately modeling liquidity constraints in cost-benefit analyses of Medicaid and similar programs

Our empirical application shows that in the hand-to-mouth model, efficient moral hazard is quite large and meaningfully changes our normative interpretation of results. The hand-to-mouth model is a reasonable model for many low-income individuals, but further work could examine both the role of high borrowing costs as well as behavioral biases, such as present-focus or overconfidence.

Our model changes the interpretation of moral hazard. Our empirical application considers access to Medicaid– a move from uninsurance to insurance that requires virtually no cost-sharing. Our results could also be applied to examine the optimal *level* of insurance in different markets,

trading off the costs (and benefits) of additional moral hazard against the risk protection gains. Accounting for liquidity constraints is also necessary to determine the optimal level of cost-sharing in many different health insurance programs, and our approach will generally imply a different level of optimal coverage for people with liquidity constraints.

Finally, while prior work on liquidity and moral hazard has focused on unemployment insurance (see Chetty, 2008) and we have extended to health insurance, liquidity constraints are likely important and affect the interpretation of moral hazard in many other insurance domains, such as auto insurance (see Cummins and Tennyson, 1996) and workers' compensation insurance (see Dionne and St-Michel, 1991). Combining data on the liquidity constraints that individuals face with models of insurance in these domains could offer new insights for their optimal design.

References

Aguiar, M., Bils, M., and Boar, C. (2024). Who are the hand-to-mouth? *Review of Economic Studies*. forthcoming.

Aron-Dine, A., Einav, L., Finkelstein, A., and Cullen, M. (2015). Moral hazard in health insurance: do dynamic incentives matter? *Review of Economics and Statistics*, 97(4):725–741.

Besley, T. J. (1988). Optimal reimbursement health insurance and the theory of Ramsey taxation. *Journal of Health Economics*, 7(4):321–336.

Brockett, P. L. and Golden, L. L. (1987). A class of utility functions containing all the common utility functions. *Management Science*, 33(8):955–964.

Brot-Goldberg, Z. C., Chandra, A., Handel, B. R., and Kolstad, J. T. (2017). What does a deductible do? The impact of cost-sharing on health care prices, quantities, and spending dynamics. *The Quarterly Journal of Economics*, 132(3):1261–1318.

Chetty, R. (2008). Moral hazard versus liquidity and optimal unemployment insurance. *Journal of Political Economy*, 116(2):173–234.

Chetty, R. and Finkelstein, A. (2013). Social insurance: Connecting theory to data. In Auerbach, A. J., Chetty, R., Feldstein, M., and Saez, E., editors, *Handbook of Public Economics*, volume 5, pages 111–193. Elsevier.

Cummins, J. D. and Tennyson, S. (1996). Moral hazard in insurance claiming: Evidence from automobile insurance. *Journal of Risk and Uncertainty*, 12:29–50.

De Meza, D. (1983). Health insurance and the demand for medical care. *Journal of Health Economics*, 2(1):47–54.

Dionne, G. and St-Michel, P. (1991). Workers' compensation and moral hazard. *The Review of Economics and Statistics*, 73:236–244.

Einav, L. and Finkelstein, A. (2018). Moral hazard in health insurance: What we know and how we know it. *Journal of the European Economic Association*, 16(4):957–982.

Einav, L., Finkelstein, A., Ryan, S. P., Schrimpf, P., and Cullen, M. R. (2013). Selection on moral hazard in health insurance. *American Economic Review*, 103(1):178–219.

Ericson, K. M. and Laibson, D. (2019). Intertemporal choice. In *Handbook of Behavioral Economics: Applications and Foundations 1*, volume 2, pages 1–67. Elsevier.

Ericson, K. M. and Sydnor, J. R. (2018). Liquidity constraints and the value of insurance. Technical report, National Bureau of Economic Research.

Finkelstein, A., Hendren, N., and Luttmer, E. F. (2019). The value of medicaid: Interpreting results from the oregon health insurance experiment. *Journal of Political Economy*, 127(6):2836–2874.

Gross, T., Layton, T. J., and Prinz, D. (2022). The liquidity sensitivity of healthcare consumption: Evidence from social security payments. *American Economic Review: Insights*, 4(2):175–190.

Kaplan, G., Violante, G. L., and Weidner, J. (2014). The wealthy hand-to-mouth. *Brookings Papers on Economic Activity*, 2014(1):77–138.

Lee, S. C. and Maxted, P. (2023). Credit card borrowing in heterogeneous-agent models: Reconciling theory and data. Technical report, Mimeo.

Marone, V. R. and Sabety, A. (2022). When should there be vertical choice in health insurance markets? *American Economic Review*, 112(1):304–342.

Mukherjee, A., Sacks, D., and Yoo, H. (2024). Does health insurance reduce consumption risk? Evidence from medicaid expansions. Technical report, Mimeo.

Nyman, J. A. (1999a). The economics of moral hazard revisited. *Journal of Health Economics*, 18(6):811–824.

Nyman, J. A. (1999b). The value of health insurance: The access motive. *Journal of Health Economics*, 18(2):141–152.

Nyman, J. A., Koc, C., Dowd, B. E., McCreedy, E., and Trenz, H. M. (2018). Decomposition of moral hazard. *Journal of Health Economics*, 57:168–178.

Olafsson, A. and Pagel, M. (2018). The liquid hand-to-mouth: Evidence from personal finance management software. *The Review of Financial Studies*, 31(11):4398–4446.

Parker, J. A., Souleles, N. S., Johnson, D. S., and McClelland, R. (2013). Consumer spending and the economic stimulus payments of 2008. *American Economic Review*, 103(6):2530–2553.

Pauly, M. V. (1968). The economics of moral hazard: Comment. *American Economic Review*, 58(3):531–537.

Pauly, M. V. and Blavin, F. E. (2008). Moral hazard in insurance, value-based cost sharing, and the benefits of blissful ignorance. *Journal of Health Economics*, 27(6):1407–1417.

Rowell, D. and Connelly, L. B. (2012). A history of the term "moral hazard". *Journal of Risk and Insurance*, 79(4):1051–1075.

Sergeyev, D., Lian, C., and Gorodnichenko, Y. (2024). The economics of financial stress. *Review of Economic Studies*. forthcoming.

Zeckhauser, R. (1970). Medical insurance: A case study of the tradeoff between risk spreading and appropriate incentives. *Journal of Economic Theory*, 2(1):10–26.

Appendix A Proofs

A.1 Proof of Proposition 1

Proof. Item 1: Define $A(m) = \frac{y-\alpha\pi m}{N}$ and $B(m) = \frac{y-(1-\alpha)m-\alpha\pi m}{N}$. Then, $m_{PL}^E(\alpha)$ is implicitly defined by

$$h'(m_{PL}^E(\alpha)) - (1-\pi)\alpha u'(A(m_{PL}^E(\alpha))) - (1-\alpha(1-\pi))u'(B(m_{PL}^E(\alpha))) = 0. \tag{A1}$$

We label Equation (A1) as F and apply the implicit function theorem to obtain $\frac{\partial m_{PL}^E(\alpha)}{\partial \alpha} = -\frac{\frac{\partial F}{\partial \alpha}}{\frac{\partial F}{\partial m}}$. By assumption it holds that $u'' < 0$ and $h'' < 0$. Thus from

$$\frac{\partial F}{\partial m} = h''(m_{PL}^E(\alpha)) + \frac{1-\pi}{N}\alpha^2\pi u''(A(m_{PL}^E(\alpha))) + \frac{(1-\alpha(1-\pi))^2}{N}u''(B(m_{PL}^E(\alpha))) < 0 \tag{A2}$$

we know that $sgn\left(\frac{\partial m_{PL}^E(\alpha)}{\partial \alpha}\right) = sgn\left(\frac{\partial F}{\partial \alpha}\right)$ where $sgn(\cdot)$ denotes the sign of an expression. Taking the partial derivative and rearranging renders

$$\begin{aligned}\frac{\partial F}{\partial \alpha} =& (1-\pi)[u'(B(m_{PL}^E(\alpha))) - u'(A(m_{PL}^E(\alpha)))] - (1-\pi)(1-\alpha)\frac{m}{N}u''(B(m_{PL}^E(\alpha))) \\ &+ (1-\pi)\pi\alpha\frac{m}{N}[u''(A(m_{PL}^E(\alpha))) - u''(B(m_{PL}^E(\alpha)))].\end{aligned} \tag{A3}$$

The first and second terms are positive from $u'' < 0$. The third term is positive if $u''' > 0$ and 0 if $u''' = 0$. This establishes $u''' \geq 0$ as a sufficient condition for $\frac{\partial m_{PL}^E(\alpha)}{\partial \alpha} > 0$.

Item 2: We abbreviate $EU_{PL}(\alpha, m_{PL}^E(\alpha))$ as EU_{PL}^E for legibility. Note that $\pi F = \frac{\partial EU_{PL}^E}{\partial m}$. Thus, by Equation (A2), the second order condition for $m_{PL}^E(\alpha)$ is fulfilled. Further, $m_{PL}^*(\alpha)$ is implicitly defined by

$$h'(m_{PL}^*(\alpha)) - (1-\alpha)u'(B(m_{PL}^*(\alpha))) = 0. \tag{A4}$$

We evaluate $\frac{\partial EU_{PL}^E}{\partial m}$ at $m_{PL}^*(\alpha)$ and obtain:

$$\left.\frac{\partial EU_{PL}^E}{\partial m}\right|_{m_{PL}^*(\alpha)} = -\pi(1-\pi)u'(A(m_{PL}^*(\alpha))) - \pi^2\alpha u'(B(m_{PL}^*(\alpha))) < 0. \tag{A5}$$

It follows that $m_{PL}^*(\alpha) > m_{PL}^E(\alpha)\ \forall\ \alpha \in]0,1]$.

Item 3: Considering Equation (A4), $m_{PL}^*(0)$ is defined by $h'(m_{PL}^*(0)) - u'(B(m_{PL}^*(0)) = 0$. When setting $\pi = 1$, Equation (A1) becomes $h'(m_{PL}^E(\alpha)) - u'(B(m_{PL}^E(\alpha))) = 0$. Thus, both $m_{PL}^*(0)$ and $m_{PL}^E(\alpha)$ are defined equivalently for $\pi = 1$, implying $m_{PL}^*(0) = m_{PL}^E(\alpha)\ \forall\ \alpha \in]0,1]$ when sickness is certain.

Item 4: Setting $\alpha = 0$ in Equation (A1) shows that $m_{PL}^E(0) = m_{PL}^*(0)$. Item 4 then follows from item 1. □

A.2 Proof of Proposition 2

Proof. Item 1: From Proposition 1 we know that under perfect liquidity and $\pi = 1$, it holds that $m^*_{PL}(0) = m^E_{PL}(\alpha)\ \forall\ \alpha \in]0,1]$. We can thus express $V^{MH}_{PL}(\alpha)$ as

$$V^{MH}_{PL}(\alpha) = EU_{PL}(\alpha, m^*_{PL}(\alpha)) - EU_{PL}(\alpha, m^E_{PL}(\alpha)). \tag{A6}$$

Further, $m^E_{PL}(\alpha)$ is the level of m that at coverage level α maximizes expected utility if $p = \pi\alpha m$. Since we know from Proposition 1 that $m^*_{PL}(\alpha) \neq m^E_{PL}(\alpha)$, $EU_{PL}(\alpha, m^*_{PL}(\alpha)) < EU_{PL}(\alpha, m^E_{PL}(\alpha))$ and thus $V^{MH}_{PL}(\alpha) < 0$. Further, because for $\pi = 1$, $EU_{PL}(\alpha, m^*_{PL}(0)) = Nu\left(\frac{y-(1-\alpha)m^*_{PL}(0)-\alpha m^*_{PL}(0)}{N}\right) + h(m^*_{PL}(0)) = Nu\left(\frac{y-m^*_{PL}(0)}{N}\right) + h(m^*_{PL}(0)) = EU(0, m^*_{PL}(0))$ it follows that $V^{RF}_{PL}(\alpha) = 0$. Thus $V_{PL}(\alpha) < 0$.

Item 2: We rearrange $V^{RF}_{PL}(\alpha)$ and utilize $u'' < 0$ to obtain

$$\begin{aligned} V^{RF}_{PL}(\alpha) &= \pi N\left[u\left(\frac{y-(1-\alpha(1-\pi))m^*_{PL}(0)}{N}\right) - u\left(\frac{y-m^*_{PL}(0)}{N}\right)\right] \\ &\quad + (1-\pi)N\left[u\left(\frac{y-\alpha\pi m^*_{PL}(0)}{N}\right) - u\left(\frac{y}{N}\right)\right] \qquad \text{(A7)} \\ &> m^*_{PL}(0)\alpha\pi(1-\pi)N\left[u'\left(\frac{y-(1-\alpha(1-\pi))m^*_{PL}(0)}{N}\right) - u'\left(\frac{y-\alpha\pi m^*_{PL}(0)}{N}\right)\right] > 0 \qquad \text{(A8)} \end{aligned}$$

Item 3: From Proposition 1 we know that $u''' \geq 0$ is sufficient for $m^E_{PL}(\alpha) > m^*_{PL}(0)\ \forall\ \alpha \in]0,1]$ when $\pi \in]0,1[$. Because $m^E_{PL}(\alpha)$ maximizes expected utility for coverage level α, it follows that $EU_{PL}(\alpha, m^E_{PL}(\alpha)) > EU_{PL}(\alpha, m^*_{PL}(0))$ and thus $V^{MHE}_{PL}(\alpha) > 0$.

Item 4: From Equation (5), we can see that $V^{MH} = 0$ if $\alpha = 0$. Abbreviating $\pi\alpha m^*_{PL}(\alpha)$ as p_α and $\pi\alpha m^*_{PL}(0)$ as p_0, we have

$$\begin{aligned} \frac{\partial V^{MH}}{\partial\alpha} =& \pi\left[\left(m^*_{PL}(\alpha) - \frac{\partial p_\alpha}{\partial\alpha} - \frac{\partial m^*_{PL}(\alpha)}{\partial\alpha}(1-\alpha)\right)u'(B(m^*_{PL}(\alpha))) + \frac{\partial m^*_{PL}(\alpha)}{\partial\alpha}h'(m^*_{PL}(\alpha))\right] \\ &- (1-\pi)\frac{\partial p_\alpha}{\partial\alpha}u'(A(m^*_{PL}(\alpha))) - \pi\left(m^*_{PL}(0) - \frac{\partial p_0}{\partial\alpha}\right)u'(B(m^*_{PL}(0))) \\ &- (1-\pi)\frac{\partial p_0}{\partial\alpha}u'(A(m^*_{PL}(0))) \end{aligned} \tag{A9}$$

Substituting the first order condition for $m^*_{PL}(\alpha)$, we obtain

$$\begin{aligned} \frac{\partial V^{MH}}{\partial\alpha} =& \pi\left(m^*_{PL}(\alpha) - \frac{\partial p_\alpha}{\partial\alpha}\right)u'(B(m^*_{PL}(\alpha))) - (1-\pi)\frac{\partial p_\alpha}{\partial\alpha}u'(A(m^*_{PL}(\alpha))) \\ &- \pi\left(m^*_{PL}(0) - \frac{\partial p_0}{\partial\alpha}\right)u'(B(m^*_{PL}(0))) - (1-\pi)\frac{\partial p_0}{\partial\alpha}u'(A(m^*_{PL}(0))) \end{aligned} \tag{A10}$$

At $\alpha = 0$, $\frac{\partial p_\alpha}{\partial \alpha} = 0\frac{\partial m^*(\alpha)}{\partial \alpha}\pi + m^*(0)\pi = \frac{\partial p_0}{\partial \alpha}$. Thus, $\left.\frac{\partial V^{MH}}{\partial \alpha}\right|_{\alpha=0} = 0$. Further

$$\begin{aligned}
\frac{\partial^2 V^{MH}}{\partial \alpha^2} =& \pi\left(\frac{\partial m^*_{PL}(\alpha)}{\partial \alpha} - \frac{\partial^2 p_\alpha}{\partial \alpha^2}\right) u'(B(m^*_{PL}(\alpha))) \\
&- \pi\left(m^*_{PL}(\alpha) - \frac{\partial p_\alpha}{\partial \alpha}\right)\left(\frac{\partial m^*_{PL}(\alpha)}{\partial \alpha}(1-\alpha) - m^*(\alpha) + \frac{\partial p_\alpha}{\partial \alpha}\right)\frac{1}{N}u''(B(m^*_{PL}(\alpha))) \\
&- (1-\pi)\left[\frac{\partial^2 p_\alpha}{\partial \alpha^2}u'(A(m^*_{PL}(\alpha))) - \left(\frac{\partial p_\alpha}{\partial \alpha}\right)^2\frac{1}{N}u''(A(m^*_{PL}(\alpha)))\right] \\
&+ \pi\left[\frac{\partial^2 p_0}{\partial \alpha^2}u'(B(m^*_{PL}(0))) - \left(m^*_{PL}(0) - \frac{\partial p_0}{\partial \alpha}\right)^2\frac{1}{N}u''(B(m^*_{PL}(0)))\right] \\
&- (1-\pi)\left[\frac{\partial^2 p_0}{\partial \alpha^2}u'(A(m^*_{PL}(0))) - \left(m^*_{PL}(0) - \frac{\partial p_0}{\partial \alpha}\right)^2\frac{1}{N}u''(A(m^*_{PL}(0)))\right] \qquad \text{(A11)}
\end{aligned}$$

Observe $\frac{\partial^2 p_\alpha}{\partial \alpha^2} = 2\frac{\partial m^*_{PL}(\alpha)}{\partial \alpha}\pi + \alpha\frac{\partial^2 m^*_{PL}(\alpha)}{\partial \alpha^2}\pi$ and $\frac{\partial^2 p_0}{\partial \alpha^2} = 0$. Evaluating at $\alpha = 0$ and again using that in this case $\frac{\partial p_\alpha}{\partial \alpha} = \frac{\partial p_0}{\partial \alpha}$ we can see that

$$\begin{aligned}
\left.\frac{\partial^2 V^{MH}}{\partial \alpha^2}\right|_{\alpha=0} =& \pi\frac{\partial m^*_{PL}(\alpha)}{\partial \alpha}\left[(1-2\pi)u'(B(m^*_{PL}(0))) - m*_{PL}(0)(1-\pi)\frac{1}{N}u''(B(m^*_{PL}(0)))\right] \\
&- \pi\frac{\partial m^*_{PL}(\alpha)}{\partial \alpha}2(1-\pi)u'(A(m^*_{PL}(0))) \qquad \text{(A12)}
\end{aligned}$$

If V^{MH} is zero with slope zero at $\alpha = 0$, then it must be positive for some values of α to the right of 0 if it is convex at $\alpha = 0$. From Equation (A4) and the implicit function theorem, we can see that

$$\frac{\partial m^*_{PL}(\alpha)}{\partial \alpha} = -\frac{u'(B(m^*_{PL}(\alpha))) - (1-\alpha)\frac{m}{N}u''(B(m^*_{PL}(\alpha)))}{h''(m^*_{PL}(\alpha)) + \frac{(1-\alpha)^2}{N}u''(B(m^*_{PL}(\alpha)))} > 0. \qquad \text{(A13)}$$

Thus, $\left.\frac{\partial^2 V^{MH}}{\partial \alpha^2}\right|_{\alpha=0}$ is positive if Condition (21) is fulfilled.

Item 5: Follows immediately from $V_{PL} = V^{RF}_{PL} + V^{MH}_{PL}$ and items 2 and 3. □

A.3 Proof of Proposition 3

Proof. Item 1: For the hand-to-mouth individuals we abbreviate $\frac{y-\alpha\pi m}{N} = A(m)$ and $\frac{y-\alpha\pi m}{N} - \frac{(1-\alpha)m}{K} = B(m)$.[21] We further abbreviate $EU_{HTM}(\alpha, m^E_{HTM}(\alpha))$ as EU^E_{HTM} for legibility. To find the efficient level of medical spending, we maximize the expected utility of the individual, including the influence of medical spending on the insurance premium. The first order condition for this level

[21] We do not index for the type of individual for legibility unless it is necessary.

is given by

$$\begin{aligned}\frac{\partial EU^E_{HTM}}{\partial m} =&\pi h'(m^E_{HTM}(\alpha)) - (1-\pi)\alpha\pi u'(A(m^E_{HTM}(\alpha)))\\ &- \pi\left[\frac{N-K}{N}\alpha\pi u'(A(m^E_{HTM}(\alpha))) + \left(\frac{\alpha\pi K}{N} + (1-\alpha)\right)u'(B(m^E_{HTM}(\alpha)))\right] = 0.\end{aligned} \tag{A14}$$

The second order condition is fulfilled as can be seen by

$$\frac{\partial^2 EU^E_{HTM}}{\partial m^2} =\pi\left[(N-K)\left(\frac{\alpha\pi}{N}\right)^2 u''(A(m^E_{HTM}(\alpha))) + K\left(\frac{\alpha\pi}{N} + \frac{1-\alpha}{K}\right)^2 u''(B(m^E_{HTM}(\alpha)))\right] \tag{A15}$$

$$+ \pi h''(m^E_{HTM}(\alpha)) + (1-\pi)\frac{(\alpha\pi)^2}{N}u''(A(m^E_{HTM}(\alpha))) < 0.$$

We now evaluate $\frac{\partial EU^E_{HTM}}{\partial m}$ at $m^*_{HTM}(\alpha)$ which is implicitly defined by

$$h'(m^*_{HTM}(\alpha)) - (1-\alpha)u'(B(m^*_{HTM}(\alpha))) = 0 \tag{A16}$$

and obtain

$$\begin{aligned}\left.\frac{\partial EU^E_{HTM}}{\partial m}\right|_{m^*_{HTM}(\alpha)} =& -\pi^2\alpha\left[\frac{N-K}{N}u'(A(m^*_{HTM}(\alpha))) + \frac{K}{N}u'(B(m^*_{HTM}(\alpha)))\right]\\ &- (1-\pi)u'(A(m^*_{HTM}(\alpha))) < 0\end{aligned} \tag{A17}$$

Because this expression is negative, $EU^E_{HTM}(\alpha)$ achieves its maximum to the left of $m^*_{HTM}(\alpha)$ and thus $m^*_{HTM}(\alpha) > m^E_{HTM}(\alpha)\ \forall\ \alpha \in]0,1]$.

Item 2: From the implicit function theorem, we can then infer that $sgn\left(\frac{\partial m^E_{HTM}(\alpha)}{\partial\alpha}\right) = sgn\left(\frac{\partial^2 EU^E_{HTM}}{\partial m\partial\alpha}\right)$. The relevant expression is

$$\begin{aligned}\frac{\partial^2 EU^E_{HTM}}{\partial m\partial\alpha} =& -\pi\left[\left(\frac{N-K}{N}\pi + (1-\pi)\right)u'(A(m^E_{HTM}(\alpha))) + \left(\frac{K}{N}\pi - 1\right)u'(B(m^E_{HTM}(\alpha)))\right]\\ &+ u''(A(m^E_{HTM}(\alpha)))m^E_{HTM}(\alpha)\pi\left[\pi^2\alpha\frac{N-K}{N^2} + (1-\pi)\pi\alpha\frac{1}{N}\right]\\ &- u''(B(m^E_{HTM}(\alpha)))m^E_{HTM}(\alpha)\pi\left[\pi\alpha\frac{N-K\pi}{N^2} + (1-\alpha)\frac{N-K\pi}{NK}\right].\end{aligned} \tag{A18}$$

The first line equals $\pi\left(1-\frac{K}{N}\pi\right)\left(u'(B(m^E_{HTM}(\alpha))) - u'(A(m^E_{HTM}(\alpha)))\right)$ and is thus positive due to $u'' < 0$. From $u''' > 0$, we know that $u''(B(m^E_{HTM}(\alpha))) < u''(A(m^E_{HTM}(\alpha)))$. For the second and third line to be positive, it thus remains to be shown that $\pi^2\alpha\frac{N-K}{N^2} + (1-\pi)\pi\alpha\frac{1}{N} \leq \pi\alpha\frac{N-K\pi}{N^2} + (1-\alpha)\frac{N-K\pi}{NK}$. This expression reduces to $0 \leq (1-\alpha)\frac{N-K\pi}{NK}$ which is true for $\alpha \in [0,1]$. Thus $u''' > 0$ is sufficient for $\frac{\partial m^E_{HTM}(\alpha)}{\partial\alpha} > 0$. For any $\alpha > 0$, it follows that $m^E_{HTM}(\alpha) > m^E_{HTM}(0) = m^*_{HTM}(0)$. $\square$

A.4 Proof of Proposition 4

The proof to this proposition follows the same steps as that for the corresponding items in the proof to Proposition 2. When the proofs differ, the steps are nevertheless stated for completeness.

Proof. Item 1: By virtue of $u'' < 0$, it follows that

$$
\begin{aligned}
V^{RF}_{HTM}(\alpha) &= EU_{HTM}(\alpha, m^*_{HTM}(0)) - EU_{HTM}(0, m^*_{HTM}(0)) && \text{(A19)}\\
&= \pi \left[(N-K) \left[u\left(\frac{y - \pi\alpha m^*_{HTM}(0)}{N} \right) - u\left(\frac{y}{N} \right) \right] \right.\\
&\quad \left. + Ku\left(\frac{y - \pi\alpha m^*_{HTM}(0)}{N} - \frac{(1-\alpha) m^*_{HTM}(0)}{K} \right) - u\left(\frac{y}{N} - \frac{m^*_{HTM}(0)}{K} \right) \right]\\
&\quad + (1-\pi)N \left[u\left(\frac{y - \pi\alpha m^*_{HTM}(0)}{N} \right) - u\left(\frac{y}{N} \right) \right] && \text{(A20)}\\
&> \alpha m^*_{HTM}(0) \left(1 - \frac{\pi K}{N} \right) \left[u'\left(\frac{y}{N} - \frac{m^*_{HTM}(0)}{K} \right) - u'\left(\frac{y - \pi\alpha m^*_{HTM}(0)}{N} \right) \right] && \text{(A21)}
\end{aligned}
$$

This expression is positive, because $u'\left(\frac{y}{N} - \frac{m^*_{HTM}(0)}{K} \right) > u'\left(\frac{y - \pi\alpha m^*_{HTM}(0)}{N} \right)$ by $u'' < 0$.

Item 2: Proof is analogous to item 3 in Proposition 2.

Item 3: As in the PL case, we can see from Equation 5 that $V^{MH}_{HTM} = 0$ if $\alpha = 0$. Using the definitions of $A(m)$ and $B(m)$ as above and abbreviating $\pi\alpha m^*_{HTM}(\alpha)$, we write

$$
\begin{aligned}
\frac{\partial V^{MH}_{HTM}}{\partial \alpha} =& -\pi \left[\frac{N-K}{N} \frac{\partial p_\alpha}{\partial \alpha} u'(A(m^*_{HTM}(\alpha))) + K \left(\frac{1}{N} \frac{\partial p_\alpha}{\partial \alpha} + \frac{1-\alpha}{K} \frac{\partial m^*_{HTM}(\alpha)}{\partial \alpha} \right.\right.\\
&\left.\left. - \frac{1}{K} m^*_{HTM}(\alpha) \right) u'(B(m^*_{HTM}(\alpha))) \right] - (1-\pi) \frac{\partial p_\alpha}{\partial \alpha} u'(A(m^*_{HTM}(\alpha)))\\
&+ \pi \frac{\partial m^*_{HTM}(\alpha)}{\partial \alpha} h'(m^*_{HTM}(\alpha)) + (1-\pi)\pi m^*_{HTM}(0) u'(A(m^*_{HTM}(0)))\\
&+ \pi m^*_{HTM} \left[\frac{N-K}{N} \pi(0) u'(A(m^*_{HTM}(0))) + K \left(\frac{\pi}{N} - \frac{1}{K} \right) u'(B(m^*_{HTM}(0))) \right] && \text{(A22)}
\end{aligned}
$$

We substitute the first-order condition for $m^*_{HTM}(\alpha)$ and obtain:

$$
\begin{aligned}
\frac{\partial V^{MH}_{HTM}}{\partial \alpha} =& -\pi \left[\frac{N-K}{N} \frac{\partial p_\alpha}{\partial \alpha} u'(A(m^*_{HTM}(\alpha))) + K \left(\frac{1}{N} \frac{\partial p_\alpha}{\partial \alpha} - \frac{1}{K} m^*_{HTM}(\alpha) \right) u'(B(m^*_{HTM}(\alpha))) \right]\\
&- (1-\pi) \frac{\partial p_\alpha}{\partial \alpha} u'(A(m^*_{HTM}(\alpha))) + (1-\pi)\pi m^*_{HTM}(0) u'(A(m^*_{HTM}(0)))\\
&+ \pi m^*_{HTM} \left[\frac{N-K}{N} \pi(0) u'(A(m^*_{HTM}(0))) + K \left(\frac{\pi}{N} - \frac{1}{K} \right) u'(B(m^*_{HTM}(0))) \right] && \text{(A23)}
\end{aligned}
$$

At $\alpha = 0$, $\frac{\partial p_\alpha}{\partial \alpha} = \pi m^*_{HTM}(0)$ and thus $\left.\frac{\partial V^{MH}_{HTM}}{\partial \alpha}\right|_{\alpha=0} = 0$. Further

$$\begin{aligned}
\frac{\partial^2 V^{MH}_{HTM}}{\partial \alpha^2} =& \pi \frac{N-K}{N} \frac{\partial^2 p_\alpha}{\partial \alpha^2} u'(A(m^*_{HTM}(\alpha))) + \pi \frac{N-K}{N} \left(\frac{\partial p_\alpha}{\partial \alpha}\right)^2 u''(A(m^*_{HTM}(\alpha))) \\
&- \pi K \left(\frac{1}{N}\frac{\partial^2 p_\alpha}{\partial \alpha^2} - \frac{1}{K}\frac{\partial m^*_{HTM}(\alpha)}{\partial \alpha}\right) u'(B(m^*_{HTM}(\alpha))) \\
&+ \pi K \left(\frac{1}{N}\frac{\partial p_\alpha}{\partial \alpha} - \frac{m^*_{HTM}(\alpha)}{K}\right)\left(\frac{1}{N}\frac{\partial p_\alpha}{\partial \alpha} + (1-\alpha)\frac{1}{K}\frac{\partial m^*_{HTM}(\alpha)}{\partial \alpha} - \frac{m^*_{HTM}(\alpha)}{K}\right) u''(B(m^*_{HTM}(\alpha))) \\
&- (1-\pi)\frac{\partial^2 p_\alpha}{\partial \alpha^2} u'(A(m^*_{HTM}(\alpha))) + (1-\pi)\left(\frac{\partial p_\alpha}{\partial \alpha}\right)^2 \frac{1}{N} u''(A(m^*_{HTM}(\alpha))) \\
&- \pi \frac{N-K}{N^2} (\pi m^*_{HTM}(0))^2 u''(A(m^*_{HTM}(0))) - (1-\pi)\frac{1}{N}(\pi m^*_{HTM}(0))^2 u''(A(m^*_{HTM}(0))) \\
&- \pi K \left(\frac{\pi m^*_{HTM}(0)}{N} - \frac{m^*_{HTM}(0)}{K}\right)^2 u''(B(m^*_{HTM}(0))) \qquad \text{(A24)}
\end{aligned}$$

Observe that as in the PL case $\frac{\partial^2 p_\alpha}{\partial \alpha^2} = 2\frac{\partial m^*_{HTM}(\alpha)}{\partial \alpha}\pi + \alpha \frac{\partial^2 m^*_{HTM}(\alpha)}{\partial \alpha^2}\pi$. Evaluating at $\alpha = 0$ and again using that in this case $\frac{\partial p_\alpha}{\partial \alpha} = \pi m^*_{HTM}(0)$ we can see that

$$\begin{aligned}
\left.\frac{\partial^2 V^{MH}_{HTM}}{\partial \alpha^2}\right|_{\alpha=0} =& \pi \frac{\partial m^*_{HTM}(\alpha)}{\partial \alpha}\Big[(N - 2K\pi)u'(B(m^*_{HTM}(0))) - 2(N - K\pi)u'(A(m^*_{HTM}(0))) \\
&- \frac{N - \pi K}{K} u''(B(m^*_{HTM}(0)))\Big] \qquad \text{(A25)}
\end{aligned}$$

Note that from the implicit definition of $m^*_{HTM}(\alpha)$ we can show

$$\frac{\partial m^*_{HTM}(\alpha)}{\partial \alpha} = -\frac{u'(B(m^*_{HTM}(\alpha))) - (1-\alpha)\frac{m}{K}u''(B(m^*_{HTM}(\alpha)))}{h''(m^*_{HTM}(\alpha)) + \frac{(1-\alpha)^2}{K}u''(B(m^*_{HTM}(\alpha)))} > 0. \qquad \text{(A26)}$$

We can thus conclude that $\left.\frac{\partial^2 V^{MH}_{HTM}}{\partial \alpha^2}\right|_{\alpha=0} > 0$ if and only if Condition (22) holds.

Item 4: Follows immediately from $V_{HTM} = V^{RF}_{HTM} + V^{MH}_{HTM}$ and items 1 and 3. □

A.5 Proof of Proposition 5

We consider the first order condition for $m^*_{HTM}(\alpha)$ in Equation (A16) and evaluate it at $m^*_{PL}(\alpha)$ as implicitly defined in Equation (A4). We obtain

$$(1-\alpha)\left[u'\left(\frac{y - \alpha\pi m^*_{PL}(\alpha) - (1-\alpha)m^*_{PL}(\alpha)}{N}\right) - u'\left(\frac{y - \alpha\pi m^*_{PL}(\alpha)}{N} - \frac{(1-\alpha)m^*_{PL}(\alpha)}{K}\right)\right]. \qquad \text{(A27)}$$

From $\frac{y-\alpha\pi m^*_{PL}(\alpha)-(1-\alpha)m^*_{PL}}{N} > \frac{y-\alpha\pi m^*_{PL}(\alpha)}{N} - \frac{(1-\alpha)m^*_{PL}}{K}$ and $u'' < 0$, we know that the above expression is negative. The second order condition for $m^*_{HTM}(\alpha)$ is fulfilled as can be seen by

$$h''(m^*_{HTM}(\alpha)) + \frac{(1-\alpha)^2}{K}u''\left(\frac{y - \alpha\pi m^*_{HTM}(\alpha)}{N} - \frac{(1-\alpha)m^*_{HTM}}{K}\right) < 0. \qquad \text{(A28)}$$

Thus we can conclude that $m^*_{PL}(\alpha) > m^*_{HTM}(\alpha)$.

A.6 Proof of Proposition 6

We again refrain from indexing individual types for brevity. All equations below pertain to hand-to-mouth individuals. We abbreviate $\frac{y-\alpha\pi m^*(\alpha)}{N} = A(m^*)$, $\frac{y-\alpha\pi m^*(0)}{N} = A(m^0)$, $\frac{y-\alpha\pi m^*(\alpha)}{N} - \frac{(1-\alpha)m^*(\alpha)}{K} = B(m^*)$, and $\frac{y-\alpha\pi m^*(0)}{N} - \frac{(1-\alpha)m^*(0)}{K} = B(m^0)$. We consider how the critical value of $\hat{\alpha}$ changes in K. To do this, we implicitly define $\hat{\alpha}$ as $V^{MH}(\hat{\alpha}) = 0$. We then apply the implicit function theorem such that $\frac{\partial\hat{\alpha}}{\partial K} = -\frac{\partial V^{MH}}{\partial K} / \frac{\partial V^{MH}}{\partial\hat{\alpha}}$. From the proof of Proposition 4 we know that if $\hat{\alpha}$ exists in $]0,1[$, then $V^{MH}(\alpha)$ crosses the x-axis from above at $\hat{\alpha}$. As such, $\frac{\partial V^{MH}}{\partial\hat{\alpha}} < 0$. The sign of $\frac{\partial\hat{\alpha}}{\partial K}$ is thus equivalent to that of $\frac{\partial V^{MH}}{\partial K}$. Because $h(m)$ is approximately linear, we approximate it as φm.

$$\begin{aligned}\frac{\partial V^{MH}}{\partial K} =&\pi[u(B(m^*)) - u(A(m^*))] + \pi(N-K)\frac{\partial A(m^*)}{\partial K}u'(A(m^*)) + \pi K\frac{\partial B(m^*)}{\partial K}u'(B(m^*))\\ &+ \pi\varphi\frac{\partial m^*(\alpha)}{\partial K} + (1-\pi)N\frac{\partial A(m^*)}{\partial K}u'(A(m^*)) - \pi[u(B(m^0)) - u(A(m^0))]\\ &- \pi(N-K)\frac{\partial A(m^0)}{\partial K}u'(A(m^0)) - \pi K\frac{\partial B(m^0)}{\partial K}u'(B(m^0)) - \pi\varphi\frac{\partial m^*(0)}{\partial K}\\ &- (1-\pi)N\frac{\partial A(m^0)}{\partial K}u'(A(m^0)) \qquad \text{(A29)}\end{aligned}$$

From the abbreviations, we know that $\frac{\partial A(m^*)}{\partial K} = -\frac{\alpha\pi}{N}\frac{\partial m^*(\alpha)}{\partial K}$, $\frac{\partial A(m^0)}{\partial K} = -\frac{\alpha\pi}{N}\frac{\partial m^*(0)}{\partial K}$, $\frac{\partial B(m^*)}{\partial K} = -\frac{\alpha\pi}{N}\frac{\partial m^*(\alpha)}{\partial K} - \frac{(1-\alpha)\frac{\partial m^*(\alpha)}{\partial K}K-(1-\alpha)m^*(\alpha)}{K^2}$, and $\frac{\partial B(m^0)}{\partial K} = -\frac{\alpha\pi}{N}\frac{\partial m^*(0)}{\partial K} - \frac{(1-\alpha)\frac{\partial m^*(0)}{\partial K}K-(1-\alpha)m^*(0)}{K^2}$. To determine $\frac{\partial m^*(\alpha)}{\partial K}$ and $\frac{\partial m^*(0)}{\partial K}$, we rely on the implicit definitions of $m^*(\alpha)$ and $m^*(0)$ and apply the implicit function theorem. $m^*(\alpha)$ is implicitly defined by $F = \varphi - (1-\alpha)u'(B(m^*)) = 0$. Thus

$$\frac{\partial m^*(\alpha)}{\partial K} = -\frac{\frac{\partial F}{\partial K}}{\frac{\partial F}{\partial m^*(\alpha)}} = -\frac{-(1-\alpha)^2 m^*(\alpha)K^{-2}u''(B(m^*))}{(1-\alpha)^2K^{-1}u''(B(m^*))} = \frac{m^*(\alpha)}{K}. \qquad \text{(A30)}$$

Similarly

$$\frac{\partial m^*(0)}{\partial K} = -\frac{\frac{\partial F}{\partial K}}{\frac{\partial F}{\partial m^*(0)}} = -\frac{-m^*(0)K^{-2}u''(B(m^0))}{K^{-1}u''(B(m^0))} = \frac{m^*(0)}{K}. \qquad \text{(A31)}$$

This establishes $\frac{\partial A(m^*)}{\partial K} = \frac{\partial B(m^*)}{\partial K} = -\frac{\alpha\pi m^*(\alpha)}{NK}$ and $\frac{\partial A(m^0)}{\partial K} = \frac{\partial B(m^0)}{\partial K} = -\frac{\alpha\pi m^*(0)}{NK}$. We substitute these results into Equation (A29) to obtain

$$\begin{aligned}\frac{\partial V^{MH}}{\partial K} =&\pi[u(A(m^0)) - u(A(m^*))] + \pi[u(B(m^*)) - u(B(m^0))] + \frac{\pi\varphi}{K}(m^*(\alpha) - m^*(0))\\ &+ \frac{\alpha\pi(\pi(N-K)+(1-\pi)N)}{NK}(m^*(0)u'(A(m^0)) - m^*(\alpha)u'(A(m^*))\\ &+ \pi\frac{\alpha\pi}{N}(m^*(0)u'(B(m^0)) - m^*(\alpha)u'(B(m^*)). \qquad \text{(A32)}\end{aligned}$$

We substitute the implicit definition of $\hat{\alpha}$: $(\pi(N-K)+(1-\pi)N)[u(A(m^0))-u(A(m^*))]-\pi K[u(B(m^*))-u(B(m^0))]=\pi\varphi(m^*-m^0)$ to obtain

$$\begin{aligned}\frac{\partial V^{MH}}{\partial K} =&\frac{N}{K}[u(A(m^0))-u(A(m^*))]+\alpha\pi\frac{N-\pi K}{NK}[m^*(0)u'(A(m^0))-m^*(\alpha)u'(A(m^*))]\\ &+\alpha\pi\frac{\pi}{N}[m^*(0)u'(B(m^0))-m^*(\alpha)u'(B(m^*))]. \end{aligned}\tag{A33}$$

From Proposition 1, we know that $m^*(\alpha)>m^*(0)$ and thus $A(m^*)<A(m^0)$. By $u''<0$, we can ascertain that

$$\begin{aligned}\frac{\partial V^{MH}}{\partial K} <&\alpha\pi\frac{N}{NK}u'(A(m^*))[m^*(\alpha)-m^*(0)]+\alpha\pi\frac{N-\pi K}{NK}[m^*(0)u'(A(m^0))-m^*(\alpha)u'(A(m^*))]\\ &+\alpha\pi\frac{\pi}{N}[m^*(0)u'(B(m^0))-m^*(\alpha)u'(B(m^*))] \end{aligned}\tag{A34}$$

$$\begin{aligned}=&\frac{\alpha\pi}{K}\Bigg[u'(A(m^*))[m^*(\alpha)-m^*(0)]+[m^*(0)u'(A(m^0))-m^*(\alpha)u'(A(m^*))]\\ &+\frac{\pi K}{N}[m^*(0)u'(B(m^0))-m^*(\alpha)u'(B(m^*))-m^*(0)u'(A(m^0))+m^*(\alpha)u'(A(m^*))]\Bigg] \end{aligned}\tag{A35}$$

$$\begin{aligned}<&\frac{\alpha\pi}{K}\Bigg[u'(A(m^*))[m^*(\alpha)-m^*(0)]+[m^*(0)u'(A(m^*))-m^*(\alpha)u'(A(m^*))]\\ &+\frac{\pi K}{N}[m^*(0)u'(B(m^0))-m^*(\alpha)u'(B(m^*))-m^*(0)u'(A(m^0))+m^*(\alpha)u'(A(m^*)]\Bigg] \end{aligned}\tag{A36}$$

$$=\frac{\alpha\pi^2K}{NK}[m^*(0)u'(B(m^0))-m^*(\alpha)u'(B(m^*))-m^*(0)u'(A(m^0))+m^*(\alpha)u'(A(m^*))]\tag{A37}$$

$$<\frac{\alpha\pi^2K}{NK}m^*(\alpha)[u'(B(m^0))-u'(B(m^*))-u'(A(m^0))+u'(A(m^*))].\tag{A38}$$

The last inequality follows, because $B(m^0)<A(m^0)$ and thus $m^*(0)(u'(B(m^0))-u'(A(m^0)))<m^*(\alpha)(u'(B(m^0))-u'(A(m^0)))$. We know that $B(m^0)-B(m^*)=(m^*(0)-m^*(\alpha))\left(\frac{\alpha\pi}{N}+\frac{1-\alpha}{K}\right)>(m^*(0)-m^*(\alpha))\frac{\alpha\pi}{N}=A(m^0)-A(m^*)$. Thus, by $B(m^*)<A(m^*)$, $u''<0$ and $u'''\geq 0$ it follows that $u'(B(m^0))-u'(B(m^*))\leq u'(A(m^0))-u'(A(m^*))$. As a consequence, the right-hand-side of Equation (A38) is negative or zero and it follows that $\frac{\partial V^{MH}}{\partial K}<0$. Thus, as long as $\hat{\alpha}$ exists, it is decreasing in K and this includes the case of $K=N$. This concludes the proof.

A.7 Proof of Proposition 7

We abbreviate $\frac{y-\alpha\pi m}{N}$ as $A_{PL}(m)$ or $A_{HTM}(m)$, $\frac{y-\alpha\pi m-(1-\alpha)m}{N}=B_{PL}(m)$, and $\frac{y-\alpha\pi m}{N}-\frac{(1-\alpha)m}{K}=B_{HTM}(m)$. Because $h(m)$ is approximately linear, we approximate it as φm. $m_i^*(\alpha)$ is implicitly defined by $\varphi-(1-\alpha)u'(B_i(m_i^*(\alpha)))=0$. This implies $B_{PL}(m_{PL}^*(\alpha))=B_{HTM}(m_{HTM}^*(\alpha))$ and

thus $m^*_{PL}(\alpha) = \frac{\alpha\pi+(1-\alpha)\frac{N}{K}}{\alpha\pi+(1-\alpha)} m^*_{HTM}(\alpha) < \frac{N}{K} m^*_{HTM}(\alpha)$ which obviously implies $m^*_{PL}(\alpha) > m^*_{HTM}(\alpha)$ for $K < N$.

We abbreviate the premium as $p_i = \alpha\pi m^*_i(\alpha)$ and state the first order condition for the coverage level of individuals with perfect liquidity as

$$\begin{aligned}\frac{\partial EU_{PL}}{\partial \alpha} =& -\pi \frac{\partial p_{PL}}{\partial \alpha} u'(B_{PL}(m^*_{PL}(\alpha))) + \pi m^*_{PL}(\alpha) u'(B_{PL}(m^*_{PL}(\alpha))) - \pi(1-\alpha)\frac{\partial m^*_{PL}(\alpha)}{\partial \alpha} u'(B_{PL}(m^*_{PL}(\alpha))) \\ &+ \pi \frac{\partial m^*_{PL}(\alpha)}{\partial \alpha}\varphi - (1-\pi)\frac{\partial p_{PL}}{\partial \alpha} u'(A_{PL}(m^*_{PL}(\alpha))). \end{aligned} \tag{A39}$$

Substituting the implicit definition of $m^*_{PL}(\alpha)$ and rearranging renders

$$\pi u'(B_{PL}(m^*_{PL}(\alpha))) = \frac{\partial p_{PL}}{\partial \alpha} \frac{\pi u'(B_{PL}(m^*_{PL}(\alpha))) + (1-\pi)u'(A_{PL}(m^*_{PL}(\alpha)))}{m^*_{PL}(\alpha)}. \tag{A40}$$

Similarly, we consider the first order condition for hand-to-mouth individuals and substitute the implicit definition of $m^*_{HTM}(\alpha)$ to obtain

$$\begin{aligned}\frac{\partial EU_{HTM}}{\partial \alpha} =& m^*_{HTM}(\alpha)\pi u'(B_{HTM}(m^*_{HTM}(\alpha))) - \frac{\partial p_{HTM}}{\partial \alpha}\left[\pi \frac{N-K}{N} u'(A_{HTM}(m^*_{HTM}(\alpha))) \right. \\ &\left. + \pi \frac{K}{N} u'(B_{HTM}(m^*_{HTM}(\alpha))) + (1-\pi)u'(A_{HTM}(m^*_{HTM}(\alpha)))\right] \end{aligned} \tag{A41}$$

We now evaluate $\frac{\partial EU_{HTM}}{\partial \alpha}$ at α^*_{PL}. From $B_{PL}(m^*_{PL}(\alpha)) = B_{HTM}(m^*_{HTM}(\alpha))$:

$$\begin{aligned}\left.\frac{\partial EU_{HTM}}{\partial \alpha}\right|_{\alpha=\alpha^*_{PL}} =& m^*_{HTM}(\alpha)\frac{\partial p_{PL}}{\partial \alpha}\frac{\pi u'(B_{PL}(m^*_{PL}(\alpha))) + (1-\pi)u'(A_{PL}(m^*_{PL}(\alpha)))}{m^*_{PL}(\alpha)} \\ &- \frac{\partial p_{HTM}}{\partial \alpha}\left[\pi \frac{N-K}{N} u'(A_{HTM}(m^*_{HTM}(\alpha))) + \pi \frac{K}{N} u'(B_{HTM}(m^*_{HTM}(\alpha))) \right. \\ &\left. + (1-\pi)u'(A_{HTM}(m^*_{HTM}(\alpha)))\right] \end{aligned} \tag{A42}$$

The sign of this expression is equivalent to the sign of

$$\begin{aligned} F =& \frac{\frac{\partial p_{PL}}{\partial \alpha}}{m^*_{PL}(\alpha)}\pi u'(B_{PL}(m^*_{PL}(\alpha))) + (1-\pi)u'(A_{PL}(m^*_{PL}(\alpha))) \\ &- \frac{\frac{\partial p_{HTM}}{\partial \alpha}}{m^*_{HTM}(\alpha)}\left[\pi \frac{N-K}{N} u'(A_{HTM}(m^*_{HTM}(\alpha))) + \pi \frac{K}{N} u'(B_{HTM}(m^*_{HTM}(\alpha))) \right. \\ &\left. + (1-\pi)u'(A_{HTM}(m^*_{HTM}(\alpha)))\right] \end{aligned} \tag{A43}$$

We now note that $\frac{\partial p_i}{\partial \alpha} = \pi m^*_i(\alpha) + \alpha\pi \frac{\partial m^*_i(\alpha)}{\partial \alpha}$. Applying the implicit function theorem to the definitions of the optimal medical spending (noting that second order conditions are fulfilled) ren-

ders

$$\frac{\partial m^*_{PL}(\alpha)}{\partial \alpha} = \frac{Nu'(B_{PL}(m^*_{PL}(\alpha))) - (1-\alpha)m^*_{PL}(\alpha)u''(B_{PL}(m^*_{PL}(\alpha)))}{-(1-\alpha)^2 u''(B_{PL}(m^*_{PL}(\alpha)))} \tag{A44}$$

and

$$\frac{\partial m^*_{HTM}(\alpha)}{\partial \alpha} = \frac{Ku'(B_{HTM}(m^*_{HTM}(\alpha))) - (1-\alpha)m^*_{HTM}(\alpha)u''(B_{PL}(m^*_{HTM}(\alpha)))}{-(1-\alpha)^2 u''(B_{HTM}(m^*_{HTM}(\alpha)))}. \tag{A45}$$

We thus know that

$$\frac{\frac{\partial p_{PL}}{\partial \alpha}}{m^*_{PL}(\alpha)} = \pi + \alpha\pi \frac{\frac{N}{m^*_{PL}(\alpha)}u'(B_{PL}(m^*_{PL}(\alpha))) - (1-\alpha)m^*_{PL}(\alpha)u''(B_{PL}(m^*_{PL}(\alpha)))}{-(1-\alpha)^2 u''(B_{PL}(m^*_{PL}(\alpha)))}. \tag{A46}$$

Applying $m^*_{PL}(\alpha) < \frac{N}{K}m^*_{HTM}(\alpha)$ renders

$$\begin{aligned}
\frac{\frac{\partial p_{PL}}{\partial \alpha}}{m^*_{PL}(\alpha)} &> \pi + \alpha\pi \frac{\frac{N}{\frac{N}{K}m^*_{HTM}(\alpha)}u'(B_{PL}(m^*_{PL}(\alpha))) - (1-\alpha)m^*_{PL}(\alpha)u''(B_{PL}(m^*_{PL}(\alpha)))}{-(1-\alpha)^2 u''(B_{PL}(m^*_{PL}(\alpha)))} && \text{(A47)}\\
&= \pi + \alpha\pi \frac{\frac{K}{m^*_{HTM}(\alpha)}u'(B_{PL}(m^*_{PL}(\alpha))) - (1-\alpha)m^*_{PL}(\alpha)u''(B_{PL}(m^*_{PL}(\alpha)))}{-(1-\alpha)^2 u''(B_{PL}(m^*_{PL}(\alpha)))} && \text{(A48)}\\
&= \frac{\frac{\partial p_{HTM}}{\partial \alpha}}{m^*_{HTM}(\alpha)} && \text{(A49)}
\end{aligned}$$

We can apply this relationship to Equationa (A43) and see that

$$\begin{aligned}
F > {} & \frac{\frac{\partial p_{HTM}}{\partial \alpha}}{m^*_{HTM}(\alpha)}\pi u'(B_{PL}(m^*_{PL}(\alpha))) + (1-\pi)u'(A_{PL}(m^*_{PL}(\alpha)))\\
& - \frac{\frac{\partial p_{HTM}}{\partial \alpha}}{m^*_{HTM}(\alpha)}\left[\pi\frac{N-K}{N}u'(A_{HTM}(m^*_{HTM}(\alpha))) + \pi\frac{K}{N}u'(B_{HTM}(m^*_{HTM}(\alpha)))\right.\\
& \left. + (1-\pi)u'(A_{HTM}(m^*_{HTM}(\alpha)))\right] \qquad \text{(A50)}
\end{aligned}$$

The sign of the right-hand-side is equivalent to

$$\begin{aligned}
G = {} & \pi u'(B_{PL}(m^*_{PL}(\alpha))) + (1-\pi)u'(A_{PL}(m^*_{PL}(\alpha))) - \left[\pi\frac{N-K}{N}u'(A_{HTM}(m^*_{HTM}(\alpha)))\right.\\
& \left. + \pi\frac{K}{N}u'(B_{HTM}(m^*_{HTM}(\alpha))) + (1-\pi)u'(A_{HTM}(m^*_{HTM}(\alpha)))\right] \qquad \text{(A51)}\\
= {} & (1-\pi)[u'(A_{PL}(m^*_{PL}(\alpha))) - u'(A_{HTM}(m^*_{HTM}(\alpha)))]\\
& + \pi\frac{N-K}{N}[u'(B_{PL}(m^*_{PL}(\alpha))) - u'(A_{HTM}(m^*_{HTM}(\alpha)))] \qquad \text{(A52)}
\end{aligned}$$

From $m^*_{PL}(\alpha) > m^*_{HTM}(\alpha)$ and $u'' < 0$ it follows that $A_{PL}(m^*_{PL}(\alpha)) < A_{HTM}(m^*_{HTM}(\alpha))$ and thus $u'(A_{PL}(m^*_{PL}(\alpha))) > u'(A_{HTM}(m^*_{HTM}(\alpha)))$. Because $B_{PL}(m^*_{PL}(\alpha)) < A_{PL}(m^*_{PL}(\alpha))$, then also $B_{PL}(m^*_{PL}(\alpha)) < A_{HTM}(m^*_{HTM}(\alpha))$ and thus $u'(B_{PL}(m^*_{PL}(\alpha))) > u'(A_{HTM}(m^*_{HTM}(\alpha)))$. Thus

$G > 0$ and in consequence $\frac{\frac{\partial p_{PL}}{\partial \alpha}}{m^*_{PL}(\alpha)} > 0$, implying $\alpha^*_{HTM} > \alpha^*_{PL}$ if we assume an interior solution and thus second-order conditions to hold.

Appendix B Comparison of Our Efficient Spending Definition to the Lump-Sum Approach of Nyman (1999a)

Here we compare the approach of Nyman (1999a) to our approach of defining efficient spending, focusing on individuals with perfect liquidity. Nyman (1999a) defines efficient moral hazard as the difference between $m^L(\alpha)$ and $m^*(0)$. For Nyman, m^L is the amount of medical spending optimally chosen by an individual who faced the full marginal cost of spending but who received a lump-sum payment L. The lump-sum payment is designed to capture the income transfer a sick individual receives, and is equal to the amount of medical spending that an insured individual would have chosen, minus premiums paid: $L = (1-\pi)\alpha m^*(\alpha)$. In terms of our model, we can define m^L implicitly through

$$h'(m^L) - u'\left(\frac{y + (1-\pi)\alpha m^*(\alpha) - m^L}{N}\right) = 0. \tag{B1}$$

This concept is not exactly equal to our definition of efficient medical spending, which we term m^E. We exemplify this for full insurance ($\alpha = 1$) and show that the lump-sum approach's spending is higher than the efficient level ($m^L > m^E$). For $\alpha = 1$, m^E is defined by

$$h'(m^E) - u'\left(\frac{y - \pi m^E}{N}\right) = 0. \tag{B2}$$

To see that $m^L > m^E$, evaluate the first-order condition for m^E at m^L. This gives

$$u'\left(\frac{y + (1-\pi)m^*(1) - m^L}{N}\right) - u'\left(\frac{y - \pi m^L}{N}\right). \tag{B3}$$

given that the second-order condition for m^E is fulfilled, $m^L > m^E$ is equivalent to the above expression being negative. Rearranging and realizing that $u'' < 0$ renders

$$\frac{y + (1-\pi)m^*(1) - m^L}{N} > \frac{y - \pi m^L}{N} \tag{B4}$$

This is equivalent to $m^*(1) > m^L(1)$. We know this is true. $m^*(\alpha)$ is defined by $h'(m^*(\alpha)) = 0$ and because $u'\left(\frac{y+(1-\pi)\alpha m^*(\alpha)-m^L}{N}\right) > 0$, it will always be the case that $h'(m^L) > 0$ and thus $m^*(\alpha) > m^L$.

Thus, $m^L > m^E$ which means that the Nyman (1999b) definition overstates the amount of efficient moral hazard compared to our definition. This is because the cash transfer of the lump-sum insurance contract is too large. $m^*(\alpha)$ includes the price effect, which needs to be excluded to calculate the correct amount. Our definition arises naturally, because we simply assume an efficient market without any information asymmetries.

Appendix C Simplified First-Order Conditions for Intuitive Illustration

The following appendix utilizes Taylor approximations to derive linear versions of the first-order conditions used in the results of the papers. These linear approximations are not used to show any additional results, but simply illustrate the mechanisms explored in the paper using simple and thus intuitive functional forms.

C.1 Perfect Liquidity Individuals

We use the first-order conditions in Equations (A1) and (A4) to show how the three levels of medical spending, $m^*_{PL}(0)$, $m^E_{PL}(\alpha)$ and $m^*_{PL}(\alpha)$, can be approximated using linear equations.

We begin with optimal medical spending in the absence of insurance. From

$$u'\left(\frac{y - m^*_{PL}(0)}{N}\right) = h'(m^*_{PL}(0)) \tag{C1}$$

we use a first-order Taylor approximation of $u'(\cdot)$ around the base income $\frac{y}{N}$ to obtain

$$\Rightarrow u'\left(\frac{y}{N}\right) - u''\left(\frac{y}{N}\right)\frac{m^*_{PL}(0)}{N} \approx h'(m^*_{PL}(0)) \tag{C2}$$

$$\Leftrightarrow 1 + r_A\frac{m^*_{PL}(0)}{N} \approx H'(m^*_{PL}(0)). \tag{C3}$$

Here, r_A is the coefficient of absolute risk aversion and $H'(m) = h'(m)/u'\left(\frac{y}{N}\right)$ is a normalized version of the marginal utility from health. Because the left-hand side of (C3) is in monetary terms, $H'(m)$ is approximately in monetary terms, too. This is intuitive, because the marginal utility derived from medical spending is divided by that derived from monetary consumption.

For efficient medical spending of insured individuals, we again start at the first-order condition and use a first-order Taylor approximation of $u'(\cdot)$ around the base income $\frac{y}{N}$ to obtain

$$\begin{aligned}
&((1-\alpha)+\pi\alpha)\,u'\left(\frac{y-(1-\alpha)m^E_{PL}(\alpha)-\pi\alpha m^E_{PL}(\alpha)}{N}\right)\\
&+(1-\pi)\alpha u'\left(\frac{y-\pi\alpha m^E_{PL}(\alpha)}{N}\right) = h'(m^E_{PL}(\alpha)) \qquad \text{(C4)}\\
\Rightarrow\ &((1-\alpha)+\pi\alpha)\left[u'\left(\frac{y}{N}\right) - [(1-\alpha)+\pi\alpha]\,u''\left(\frac{y}{N}\right)\frac{m^E_{PL}(\alpha)}{N}\right]\\
&+(1-\pi)\alpha\left[u'\left(\frac{y}{N}\right) - \pi\alpha u''\left(\frac{y}{N}\right)\frac{m^E_{PL}(\alpha)}{N}\right] \approx h'(m^E_{PL}(\alpha)) \qquad \text{(C5)}\\
\Leftrightarrow\ &1 + (1-(1-\pi)\alpha(2-\alpha))\,r\frac{m^E_{PL}(\alpha)}{N} \approx H'(m^E_{PL}(\alpha)) \qquad \text{(C6)}
\end{aligned}$$

Comparing Equations (C3) and (C6), we can see item 4 of Proposition 1, namely that $m^E_{PL}(\alpha) > m^*_{PL}(0)$ for $\alpha > 0$. This is because $1-(1-\pi)\alpha(2-\alpha) < 1$ and we thus know that the slope for the efficient spending with insurance is lower than that in the case without insurance.

Lastly, we consider medical spending under moral hazard. We again start at the first order condition, apply the approximation and rearrange.

$$(1-\alpha)u'\left(\frac{y-(1-\alpha)m^*_{PL}-\pi\alpha m^*_{PL}}{N}\right) = h'(m^*_{PL}) \tag{C7}$$

$$\Rightarrow (1-\alpha) + \left(1-(2-\pi)\alpha+(1-\pi)\alpha^2\right) r\frac{m^*_{PL}(\alpha)}{N} \approx H'(m^*_{PL}). \tag{C8}$$

Thus, for the linearized definition of $m^*_{PL}(\alpha)$, both the intersection with the y axis and the slope is smaller than for the line defining efficient medical spending. As a result, $m^*_{PL}(\alpha) > m^E_{PL}(\alpha)\ \forall\ \alpha \in]0,1], \pi \in]0,1]$, which is item 2 of Proposition 1.

C.2 Hand-to-mouth Individuals

The process here is the same as above. For optimal medical spending, we begin at the first-order condition defined in Equation (A16) and apply the first-order Taylor approximation to $u'(\cdot)$ around the base income $\frac{y}{N}$. Thus

$$u'\left(\frac{y}{N} - \frac{m^*_{HTM}(0)}{K}\right) = h'(m^*_{HTM}(0)) \tag{C9}$$

$$\Rightarrow 1 + r\frac{m^*_{HTM}(0)}{K} \approx H'(m^*_{HTM}(0)). \tag{C10}$$

Because $K < N$, we know that the slope is steeper for HTM individuals than for PL individuals. This lets us conclude that $m^*_{HTM}(0) < m^*_{PL}(0)$, as is the consequence of Proposition 5.

For the efficient medical spending of insured individuals, we follow the same process as above, using the first-order condition in Equation (A14). Abbreviating $m^E_{HTM}(\alpha)$ as m in the first two lines for legibility, we obtain

$$\begin{aligned}
&\frac{N-K}{N}\alpha\pi u'\left(\frac{y-\alpha\pi m}{N}\right) + \left(\frac{\alpha\pi K}{N}+(1-\alpha)\right)u'\left(\frac{y-\alpha\pi m}{N}-\frac{(1-\alpha)m}{K}\right)\\
&+(1-\pi)\alpha u'\left(\frac{y-\alpha\pi m}{N}\right) = h'(m)
\end{aligned} \tag{C11}$$

$$\begin{aligned}
\Rightarrow\ &\frac{N-K}{N}\alpha\pi u'\left(\frac{y}{N}\right) - \frac{N-K}{N}(\alpha\pi)^2 u''\left(\frac{y}{N}\right)\frac{m}{N} + \left(\frac{K}{N}\alpha\pi+(1-\alpha)\right)u'\left(\frac{y}{N}\right)\\
&-\left(\frac{K}{N}\alpha\pi+(1-\alpha)\right)\alpha\pi u''\left(\frac{y}{N}\right)\frac{m}{N} - \left(\frac{K}{N}\alpha\pi+(1-\alpha)\right)(1-\alpha)u''\left(\frac{y}{N}\right)\frac{m}{K}\\
&(1-\pi)\alpha u'\left(\frac{y}{N}\right) - (1-\pi)\alpha^2\pi u''\left(\frac{y}{N}\right)\frac{m}{N} \approx h'(m)
\end{aligned} \tag{C12}$$

$$\Leftrightarrow 1 + \left((2-\alpha)\alpha\pi\frac{K}{N}+(1-\alpha)^2\right) r\frac{m^E_{HTM}(\alpha)}{K} \approx H'(m^E_{HTM}(\alpha)). \tag{C13}$$

With $K < N$, it is obvious that $(2-\alpha)\alpha\pi\frac{K}{N} + (1-\alpha)^2 < 1$ and thus $m^E_{HTM}(\alpha) > m^*_{HTM}(0)$, as is described in item 2 of Proposition 3.

Lastly, we again consider medical spending under moral hazard. Using the same procedure as above renders

$$(1-\alpha)u'\left(\frac{y-\pi\alpha m^*_{HTM}(\alpha)}{N} - \frac{(1-\alpha)m}{K}\right) = h'(m^*_{HTM}(\alpha)) \tag{C14}$$

$$\Rightarrow (1-\alpha) - \left((1-\alpha)\pi\alpha\frac{K}{N} + (1-\alpha)^2\right) r\frac{m^*_{HTM}(\alpha)}{K} \approx H'(m^*_{HTM}(\alpha)) \tag{C15}$$

Interpretations are the same as for the perfect liquidity case. The intersection with the y axis is lower and the slope is smaller because $(2-\alpha)\alpha\pi\frac{K}{N} + (1-\alpha)^2 - \left((1-\alpha)\pi\alpha\frac{K}{N} + (1-\alpha)^2\right) = \alpha\pi\frac{K}{N} > 0$ for $\alpha > 0$. Thus $m^*_{HTM}(\alpha) > m^E_{HTM}(\alpha)\ \forall\ \alpha \in]0,1]$ as is stated in item 1 of Proposition 3.

C.3 Specific scenarios

The two scenarios for certain losses (that is, $\pi = 1$) and full insurance (that is, $\alpha = 1$) are particularly helpful to build intuition. We cover both of them below.

C.3.1 Certain losses: $\pi = 1$

The three medical spending levels for perfect liquidity individuals are described by

$$1 + r\frac{m^*_{PL}(0)}{N} \approx H'(m^*_{PL}(0)) \tag{C16}$$

$$1 + r\frac{m^E_{PL}(\alpha)}{N} \approx H'(m^E_{PL}(\alpha)) \tag{C17}$$

$$(1-\alpha) + (1-\alpha)\, r\frac{m^*_{PL}(\alpha)}{N} \approx H'(m^*_{PL}(\alpha)). \tag{C18}$$

We can easily see that $m^*_{PL}(0) = m^E_{PL}(\alpha)\ \forall\ \alpha$. Further, slope and intersect are lower for medical spending under moral hazard such that $m^*_{PL}(\alpha) > m^E_{PL}(\alpha)$. This implies that for perfect liquidity individuals, under certain losses there is no value of efficient moral hazard, $V^{MHE}_{PL}(\pi = 1) = 0$, while the value of inefficient moral hazard is negative, $V^{MHI}_{PL}(\pi = 1) < 0$ and thus the total moral hazard value will always be negative: $V^{MH}_{PL}(\pi = 1) < 0$.

Considering HTM individuals, we see that

$$1 + r\frac{m^*_{HTM}(0)}{K} \approx H'(m^*_{HTM}(0)) \tag{C19}$$

$$1 + \left((2-\alpha)\alpha\frac{K}{N} + (1-\alpha)^2\right) r\frac{m^E_{HTM}(\alpha)}{K} \approx H'(m^E_{HTM}(\alpha))) \tag{C20}$$

$$(1-\alpha) - \left((1-\alpha)\alpha\frac{K}{N} + (1-\alpha)^2\right) r\frac{m^*_{HTM}(\alpha)}{K} \approx H'(m^*_{HTM}(\alpha)). \tag{C21}$$

This lets us see relatively easily that $m^*_{HTM}(\alpha) > m^E_{HTM}(\alpha) > m_{HTM}(0)$. The latter relationship contrasts the HTM and PL cases.[22] From it, it is obvious that even in the presence of certain losses, there is a positive value of efficient moral hazard for hand-to-mouth individuals, $V^{MHE}_{HTM}(\pi = 1) > 0$. Combined with the negative value of inefficient moral hazard, the sign of the total moral hazard value is undetermined for the hand-to-mouth case.

C.4 Full insurance: $\alpha = 1$

The case of full insurance is particularly informative to compare the medical spending of both types of individuals. We restate the definition of medical spending without any insurance as

$$1 + r\frac{m_{PL}(0)}{N} \approx H'(m_{PL}(0)) \tag{C22}$$

$$1 + r\frac{m_{HTM}(0)}{K} \approx H'(m_{PL}(0)). \tag{C23}$$

Because of $K < N$, we know that $m_{PL}(0) > m_{HTM}(0)$.

Under moral hazard, both types will show the same spending with full insurance

$$0 = H'(m^*_{PL}(\alpha)) = H'(m^*_{HTM}(\alpha)). \tag{C24}$$

This result obtains, because full insurance makes hand-to-mouth individuals act as if they had perfect liquidity. Thus, both types also have the same efficient medical spending as can be seen by

$$1 + \pi r\frac{m^E_{PL}(\alpha)}{N} \approx H'(m^E_{PL}(\alpha)) \tag{C25}$$

$$1 + \pi r\frac{m^E_{HTM}(\alpha)}{N} \approx H'(m^E_{HTM}(\alpha)). \tag{C26}$$

For full insurance, the two types only differ in their counterfactual of no insurance. This difference has implications for the difference in the values of efficient moral hazard. Because all other levels of medical spending are the same, we know that $V^{MHE}_{HTM}(\alpha = 1) > V^{MHE}_{PL}(\alpha = 1)$ and thus $V^{MH}_{HTM}(\alpha = 1) > V^{MH}_{PL}(\alpha = 1)$. This illustrates Corollary 1.

[22] We can see this from $(2 - \alpha)\alpha\frac{K}{N} + (1 - \alpha)^2 < (2 - \alpha)\alpha + (1 - \alpha)^2$. The latter expression is 1 at $\alpha = 1$ and increasing in α so will be smaller 1 as α decreases.

Appendix D Insurance Coverage and Demand Elasticity

Starting with Zeckhauser (1970) and formalized by Besley (1988), optimal health insurance coverage has been connected to the price elasticity of demand for medical care. Our model allows for a similar connection and we can see how the connection differs for hand-to-mouth individuals in comparison to individuals with perfect liquidity.

We abbreviate $\frac{y-\alpha\pi m}{N}$ as $A_{PL}(m)$ or $A_{HTM}(m)$, $\frac{y-\alpha\pi m-(1-\alpha)m}{N} = B_{PL}(m)$, and $\frac{y-\alpha\pi m}{N} - \frac{(1-\alpha)m}{K} = B_{HTM}(m)$. Starting with individuals with perfect liquidity, we can form the first-order condition for the optimal level of coinsurance and substitute the first-order condition for optimal medical spending to have

$$\pi m^*_{PL}(\alpha^*)u'(B_{PL}) = \frac{\partial p_{PL}}{\partial \alpha} N \left(\pi u'(B_{PL}) + (1-\pi)u'(A_{PL})\right). \tag{D1}$$

Here $p_{PL} = \frac{\pi\alpha m^*(\alpha)}{N}$ is the premium for individuals with perfect liquidity. From $\frac{\partial p_{PL}}{\partial \alpha} = \frac{\pi}{N} m^*_{PL}(\alpha) + \frac{\pi}{N}\alpha\frac{\partial m^*_{PL}(\alpha^*)}{\partial \alpha}$, we rearrange the above to obtain

$$1 + \frac{\alpha^*}{m^*_{PL}(\alpha^*)}\frac{\partial m^*_{PL}(\alpha^*)}{\partial \alpha} = \frac{u'(B_{PL})}{\pi u'(B_{PL}) + (1-\pi)u'(A_{PL})}. \tag{D2}$$

Note that $\frac{\alpha^*}{m^*_{PL}(\alpha^*)}\frac{\partial m^*_{PL}(\alpha^*)}{\partial \alpha}$ is the alpha elasticity of demand for medical care, which we denote e^{α}_{PL}. Because we have normalized the price of medical care to 1 for uninsured individuals, it is $1-\alpha$ when people have insurance. e^{α}_{PL} is thus an inverse version of the price-elasticity of demand for medical care.

The equation

$$1 + e^{\alpha}_{PL} = \frac{u'(B_{PL})}{\pi u'(B_{PL}) + (1-\pi)u'(A_{PL})} \tag{D3}$$

lets us make similar conclusions as Besley (1988). When demand is perfectly unelastic and $e^{\alpha}_{PL} = 0$, then the right-hand-side must be equal one, which implies $A_{PL} = B_{PL}$ and thus $\alpha = 1$ and full insurance. As the demand becomes more elastic, e^{α}_{PL} will become positive and the fraction on the right-hand-side has to be larger than one, which is the case when there is only partial insurance.

For hand-to-mouth individuals, the relationship looks similar as can be seen by

$$1 + e^{\alpha}_{HTM} = \frac{u'(B_{HTM})}{\pi\frac{N-K}{N}u'(A_{HTM}) + \pi\frac{K}{N}u'(B_{HTM}) + (1-\pi)u'(A_{HTM})}. \tag{D4}$$

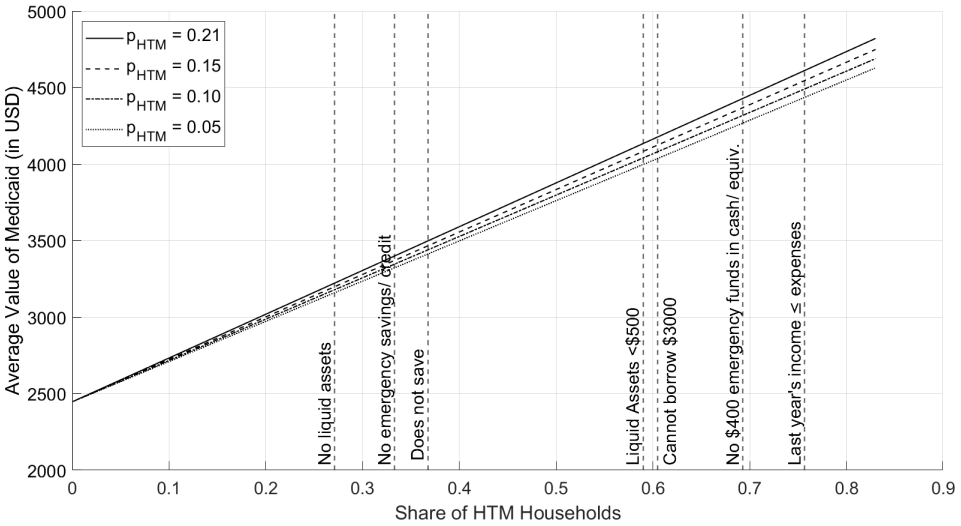
The difference between the two types of individuals lies in the magnitude of the reaction of the optimal coinsurance level. For a given measured value of e^{α}, the right-hand side of Equation (D4) dictates a higher level of coinsurance than that of Equation (D3). This mirrors our result from Proposition 7.

Appendix E Robustness: Alternative Uninsured Prices

Our main calibration for new estimates of the value of Medicaid in Section 4 assumes that all individuals, whether hand-to-mouth or with perfect liquidity, face the same marginal price of medical care when uninsured. For this, we use the empirical average of 0.21. However, it is possible that this price is different for the different types of individuals.

To explore this, we repeat the analysis allowing for different marginal prices for both groups. We assume three marginal prices lower than 0.21 for the hand-to-mouth individuals, namely 0.15, 0.1 and 0.05. Note, however, that the empirical average still has to hold for the overall population. Thus, if the share of hand-to-mouth individuals is ζ and their price is p_{HTM}, then the price for individuals with perfect liquidity needs to be $p_{PL} = \frac{0.21-\zeta p_{HTM}}{1-\zeta}$. Given a value for p_{HTM}, this limits the maximum share of hand-to-mouth individuals to some number smaller than 1, if we impose that $p_{PL} \leq 1$. Because we do not know who in the data of FHL is liquidity constrained, we calculate the average value of Medicaid for the entire population with both prices and calculate the final average population value according to $\gamma = \zeta\gamma(p_{HTM}] + (1-\zeta)\gamma(p_{PL})$.

Figure E1 – Alternative Uninsured Price Assumptions for Hand-to-mouth Individuals

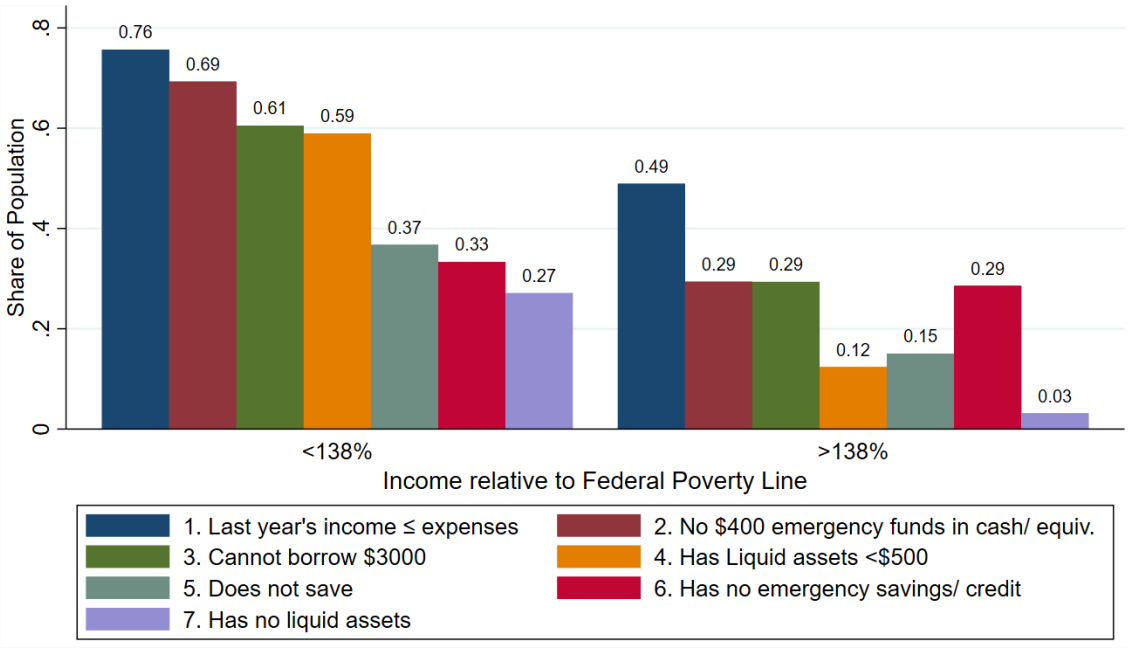


Note: The figure displays the value of γ, the estimated willingness to pay for Medicaid, for possible shares (ζ) of hand-to-mouth ($K = 1$) households in the population when those households face a marginal price of medical care of p_{HTM}. All other households are assumed to have perfect liquidity ($K = 12$) and face price $p_{PL} = \frac{0.21-\zeta p_{HTM}}{1-\zeta}$. The vertical lines show the values of potential indicators for this share from the Survey of Consumer Finances and the Survey of Household Economics and Decision-making as they are defined in Appendix F.

Results in Figure E1 show that changing p_{HTM} to a value lower than 0.21 has very little influence on the average value of Medicaid. While this value decreases for hand-to-mouth individuals, the value increases for those with perfect liquidity, such that the effect is almost completely offset, even if we assume a low value such as $p_{HTM} = 0.05$.

Appendix F Proxies for Hand-to-Mouth Status

Figure F1 – Hand-to-Mouth Indicators by Income Relative to the Federal Poverty Line



Note: Data are from a pooled analysis of the 2013, 2016, 2019 and 2022 waves of the Survey of Consumer Finances (SCF, N = 85,785) except for the question on having emergency savings or credit, which was only asked in 2022, 2019, and 2016 (N = 53,267). Data on having $400 in emergency funds which is taken from the 2023 wave of the Survey of Household Economics and Decisionmaking (SHED, N = 8,390).

It is difficult to get estimates of the fraction of people who are liquidity constrained, and more explicitly, who act as if they are hand-to-mouth. We therefore get a range of empirical proxies. To do so, we turn to two different surveys. First, we use the Survey of Consumer Finances (SCF), pooling across the 2013, 2016, 2019 and 2022 waves, totaling N=17,157. We supplement this with the data on having $400 in emergency funds which is taken from the 2023 wave of the Survey of Household Economics and Decisionmaking (SHED, N = 8,390). We want to describe the experience of low-income individuals who are most likely to be eligible for Medicaid. We therefore split the sample based on whether the individual is in a household that is above or below 138% of the Federal Poverty Level. Data on the federal poverty line for the respective years is taken from the Poverty Guidelines for 48 Contiguous States provided by the Office of the Assistant Secretary for Planning and Evaluation.[23]

Figure F1 displays the fraction of the sample displaying each of the following criteria:

[23] In the SHED, a household is categorized as being below 138% of the poverty line, if it reports an income in a category where the lower bound is below this threshold. This slightly understates the prevalence of liquidity indicators for poorer households.

1. Has zero or negative savings in the last calendar year (in blue)
2. Would not pay an unexpected $400 expense in cash or equivalent (in maroon),
3. Answered negatively to the question "In an emergency could you (or your {husband/wife/partner}) get financial assistance of $3,000 or more from any friends or relatives who do not live with you?" (in green)
4. Has liquid assets less than $500, defined as the sum of reported (market) values of all checking accounts, saving accounts, certificates of deposit, money market accounts, mutual funds, bonds, and stocks owned by the household (in yellow),
5. Answers "Don't save" to the question "Which of the following best describes your saving habits?" (in gray),
6. Answers "Postpone bills", "get an extra job", or "other" to the question "If tomorrow you experienced a financial emergency that left you unable to pay all of your bills, how would you deal with it? Would you borrow money, would you spend out of savings or investments, would you postpone paying bills, work more or get an extra job, or would you do something else?" (in red),
7. Has no liquid assets, defined as the sum of reported (market) values of all checking accounts, saving accounts, certificates of deposit, money market accounts, mutual funds, bonds, and stocks owned by the household (in lavender)