Estimating the Average Marginal Utility of Income From Willingness to Pay Conditional on Ability to Pay

Nathan Mather*

October 12, 2022

Click here for latest version

Abstract

This paper empirically estimates the degree to which the marginal utility of income changes across income groups. The estimation is based on survey responses indicating willingness to pay to avoid unpleasant experiences and relies on the assumption that the associated discomfort, or disutility, is equal on average across income groups. This assumption implies that any differences in average willingness to pay for relief are entirely driven by differences in marginal utilities of income. The results suggest that marginal utility of income is constant across income groups, implying that (cardinal) utility is roughly linear in dollars.

*Department of Economics University of Michigan: njmather@umich.edu

Thank you to my comittee members,James Hines, Ash Craig, Joel Slemrod, and Tanya Rosenblat for thier comments and support on this project.

1 Introduction

What exactly is utility and is it comparable across people? Is it measurable? Despite playing a ubiquitous role in economic theory, the definition of utility and it's use within that theory varies. Economic work generally models individual behavior as the result of utility maximization. If utility is defined as ordinal numbers corresponding to the preferences ranking of an individual than utility maximization is tautologically true so long as that person's preferences are consistent and rational. Rather than being a shortcoming, this tautology is a great strength of ordinal utility as a flexible and adaptable modeling tool. It doesn't matter *why* people choose something, if they do in fact choose it consistently, we can assign it more ordinal utility.

Normative economics research and tools frequently require the additional assumption that utility not only order individual actions, but also corresponds to an individual's welfare or well-being. This need not be happiness or pleasure per se, but in these cases utility must be something that a policymaker would conceivably want to increase. Using this stricter interpretation of utility, I estimate the degree to which the marginal utility of income changes across income groups by using survey responses indicating willingness to pay to avoid carefully chosen unpleasant experiences. The key identifying assumption is that the associated discomfort, or disutility, from these experiences is equal on average across income groups. I argue this is likely satisfied for a wide range of ethical theories of utility and well-being. This assumption implies that any differences in average willingness to pay for relief are entirely driven by differences in marginal utilities of income. The results suggest that marginal utility of income is constant across income groups, implying that (cardinal) utility is roughly linear in dollars.

Measuring utility has been recognized as an important step in policy assessment as early as 1871 when Jevons proposed that "We cannot really tell the effect of any change in trade or manufacture until we can with some approach to truth express the laws of the variation of utility numerically" (Jevons, [1871] 1911). While Jevons was optimistic about the future, stating it will eventually be a "mathematical problem of no great difficulty how to disentangle the functions expressing the degrees of utility of various commodities" (Jevons, [1871] 1911), a somewhat common view among late 19^{th} and early 20^{th} century scholars, it has turned out to be much more difficult (Fisher, 1927; Moscati, 2018). Monotonic transformations of utility do not affect the marginal rate of substitution or, by extension, people's preferences and observable choices. Despite Jevon's vision of modern computing power disentangling utility from observed behavior, it is simply not a problem that can be solved without some additional assumption.

I overcome this problem by finding a particular type of good that meets the necessary assumptions for identifying the ratio of marginal utility across groups. For now, let's just call this special good a widget. Suppose we know that the welfare benefit of receiving a widget is not affected by changes in income. That is, utility from widgets and income are independent. Now if we see that median Americans are willing to pay \$X, and billionaires are willing to pay \$Y for a widget, we know that they value X and Y dollars the same as well. This means the ratio of the average utility of money between billionaires and median Americans is Y/X. This identifies how marginal utility of income changes across income groups, or, put another way, it identifies the the concavity of total utility. This does not identify the absolute value of a util or levels of utility across income groups. This also identifies the *average* marginal utility of income conditional on observable traits like income or financial health. Focusing on the marginal utility of income across a particular subset of dimensions, like income, is often accomplished in theoretical models with homogeneity assumptions, but I show in section A.2 how the conditional expectation can be useful without assumptions of homogeneity.

While this may seem surprisingly simple, the difficulty lies in finding real goods like the widget described. The marginal utility of any given good almost always depends on the quantity already consumed and/or the consumption of another set of goods. As an extreme example, suppose the taste of chocolate is equally enjoyable for everyone and, beyond that, suppose everyone shares the same cardinal utility function. This like an ideal good for the example above, but it is unlikely to satisfy the necessary assumption. The marginal utility of a chocolate bar still depends on how much chocolate I already have as well as how many other desserts or food of any kind I consume. Since relaxing the budget constraint, changing income, will change the amount and quality of chocolate and other foods I eat, the marginal utility of a chocolate bar is unlikely to stay the same across income even in this extreme example. We don't actually need the taste to be identical for everyone, but even in that exceptional case, marginal utility is unlikely to be independent from income because the marginal utility is not independent from the consumption of other goods.

I field a survey that elicits the willingness to pay for relief from common minor pains. Using these questions, the identifying assumption is that pain relief from common minor pains is a good with marginal utility that is actually independent from income. Consider the following question. Imagine you bump your shin badly on a hard edge, for example, on the edge of a glass coffee table [1]. What is the most you would pay in U.S. dollars to completely and immediately eliminate any pain caused by the situation described, as if the event never happened? What other consumption goods make bumping one's shin more or less desirable? There is nothing available to buy to immediately relieve the pain. Any medications will not

[1] Question pulled from (Ruscheweyh, Marziniak, Stumpenhorst, Reinholz, & Knecht, 2009)

be accessible or take affect before the pain has subsided naturally. The only real option to relieve the pain is basic things everyone has access to for free, such as sitting down and rubbing it. This also means that no one, regardless of income, is already consuming some quantity of the good because it doesn't actually exist. Everyone is considering a change from zero relief to total relief. I fielded a survey asking four open ended questions like the one above as well as three binary yes or no questions with a price randomly selected. The full survey can be seen in appendix E.

Surprisingly, the results indicate that the willingness to pay for relief from these situations, and by extension the marginal utility of income, does not change much across income. This means that cardinal utility is roughly linear in dollars which contradicts the law of diminishing marginal utility (at least with respect to money). Now this analysis relies on the assumption that people maximize not just some ordinal utility function, but their own well-being. While not considered a law of economics, this is still a very common assumption. The results could be interpreted as evidence that this assumption is not met. I discuss the justifications for and interpretations of both possibilities in section 8.

The rest of the paper precedes as follows. Before diving into the analysis, section 2 outlines why cardinal preferences and interpersonal comparisons would be useful or desirable. Section 3 outlines the main assumptions of the model. This covers some of the more nuanced theoretical points that are more implicit in the empirical model. Section 4 covers the previous literature on measuring utility. Section 5 outlines the empirical model used for my painful experience survey. Section 6 outlines the data from the survey, the survey method, and summary statistics about the population. Section 7 details the analysis and results from my survey while section 8 discuses the findings. Section A covers some theory extensions that support the discussion and results but are not central to the main analysis.

2 Why Meausre Utility

Why would we want to measure utility? More generally, what value does understanding the intensity of individual preferences bring? Much of economic theory is compatible with ordinal utility that simply ranks a consumer's preferences. Normative economics. however, often attempts to use preferences to make interpersonal comparisons and engage in discussions about welfare or well-being. This requires more of utility than a simple ordering of choices.

Consider the following hypothetical scenario. A city has enough money set aside in their budget to either expand public transit or widen the roads to allow more cars through at a time. Of course, which policy they *should* implement is partially a normative question with a wide range of potential considerations. Despite being a question of ethics, economics can still

objectively aid in decision making by laying out the outcomes from each policy in a way that allows policymakers to discern the option that best coincides with their ethical normative preferences. While economic work can take a general instrumental rationality approach and consider ethical concerns like racial disparities, inequality, justice, rights, or fairness, it generally focuses on consequentialist and welfarist approaches to policy assessment that rely on aggregating preferences.

Economists have tried to find a helpful way to aggregate individual ordinal preferences for social choices. Sticking to ordinal preferences has practical appeal since preference rankings are directly observable through actions and require assuming only that people have complete and transitive preferences.[2]. Arrow (1950), however, showed that aggregating ordinal preferences is not possible without the social choice mechanism having properties generally considered undesirable. This motivated even him to look for a way to think about cardinal utility and interpersonal comparisons (Arrow, 1978). Beyond arrow showing the necessity of cardinality for social choice, ordinal preferences simply do not capture all of the relevant information for many normative decisions.

Returning to the above example about expanding public transit or widening roads, suppose the only information we have is that more people prefer widening roads. With only two options Arrow's impossibility theorem does not bind and I expect many policymakers would propose going with the option more people prefer (Arrow, 1950). However, suppose the folks favoring widening the roads favor it because it shaves a few minutes off of their morning commute and they don't use public transit. Now on the other hand, if the people in favor of expanding public transit would now be able to commute to more lucrative job opportunities or save considerable time on their commutes, reasonable policymakers may change their decision in favor of public transit expansion. The level of benefits going to each person in addition to their ranking often provides use-full, relevant information[3].

The need for more information than ordinal preferences provide is consistent with a welfarist approach to policy assessment. A welfarist policymaker has some weight they place on different people's utility and summing those weighted values up will lead to the outcome that coincides with their values. In the above example, a welfarist would, in their perfect world, take the utility gain for each person under each potential outcome, multiply that utility gain by that person's corresponding welfare weight, and sum that product for each person under both policies. Whichever sum is higher, is the policy they prefer. Doing this is, of course, easier said than done. There are many potential outcomes to consider and weigh against each other. Implicitly, the information about commuting times and job

[2]Although these are not necessarily weak assumptions

[3]Pearce (2021) explained the insufficiency of ordinal preferences in a similar way that was very helpful

opportunities is conveying some information about the intensity of the benefits, but what about other concerns like carbon emissions? How can policymakers weigh the impact of various outcomes? What about situations where the particular outcomes are difficult to measure?

Economic analysis often tries to collapse these concerns into a single dimension using measures based on willingness to pay (WTP) [4]. Willingness to pay has a clear theoretical connection to an individual's decision utility since trading money for a good implies one prefers the money less than or equal to the good traded for it. This is also consistent with the economic practice of using people's own preferences to infer what increases their well-being. While the connection to an individual's decision utility is clear, how to handle interpersonal comparisons is not.

Rather than addressing the tricky issue of interpersonal comparisons, economic work often resort to efficiency arguments that use willingness to pay, but rely only on an ordinal conception of utility for aggregation. This practice stems from the idea put forth by Hicks (1940) and Kaldor (1939) who proposed measuring economic efficiency as the sum of aggregate real income. The core idea of their argument is that, of course policy creates winners and losers, but if the winners can compensate the losers, the size of the economic pie has increased and everyone can be made better off. This idea has become widely accepted in economic analysis and modern practitioners often use such arguments to justify the using the sum of consumer and producer surplus as a measure of economic efficiency. The practice is ubiquitous, but some examples include industrial organization's analysis of consumer surplus in merger litigation (Glick, 2018; Wilson, 2019), cost benefit analysis, or theoretical policy arguments for things like price gauging (Zwolinski, 2008).

While an ethics free[5] efficiency is appealing, the idea does not hold up in a setting with heterogeneous preferences. Samuelson (1950) shows how even in the case where everyone is truly made better off, an outcome cannot be said to be efficient unless it completely expands the utility possibility frontier for every heterogeneous group. The logic here is that although everyone may be better off than the status quo after implementing a policy, that policy might also make redistributing utility to a particular group more difficult. If you are a policymaker that wants to redistribute to that group, the policy is a bad idea despite the Pareto gain relative to the status quo. This can be illustrated by a policy that shifts the slope of the utility possibility frontier between two groups and creates an intersection rather than an efficient expansion. In his own words, Samuelson (1950, pg 10) says that without comparing

[4] compensated and equivalent variation or consumer surplus

[5] It's not really ethics free since Pareto efficiency is itself a fairly strong consequentialist ethical assumption, but it avoids normative interpersonal comparisons

an "Infinite number of points, no acceptable definition of an increase in potential real income can be devised at the non-ethical level of the new welfare economics."

Samuelson's critique applies even when allowing for theoretically cost-less lump sum transfers, but as he continues to explain, movement along a utility possibility frontier would require "an ideally perfect and unattainable system of absolutely lump-sum taxes or subsidies (Samuelson, 1950, pg. 18)". Lump sum transfers to increase equity simply don't exist. Rather than mapping out a utility possibility frontier, then, an efficiency approach would require mapping out utility feasibility frontiers. On this front, there has been some work. Coate (2000) lay out a theoretical foundation that Hendren (2020) follows up on. The core idea being that, rather than use lump sum transfers as the benchmark, a given policy needs to be compared to redistribution through other government levers. Hendren (2020) looks at the cost of redistribution through the income tax code as a benchmark for other policies. While this is a promising improvement on the lump sum transfer arguments, it does not absolve out need for interpersonal comparisons. First, it does not address Samuelson's concern over intersecting utility possibility frontiers[6]. More fundamentally, even a perfect efficiency measure can only partially order outcomes by eliminating Pareto dominated policies. A policymaker may still face a difficult and complex decision that would be eased by objective measures.

A slightly different version of efficiency arguments is also frequently employed in attempt to separate economic analysis from interpersonal comparisons. Introductory economics textbooks frequently present policy analysis as a balance between equity and efficiency. Betsey Stevenson and Justin Wolfers' book lays out the idea clearly

> One argument for focusing on efficiency is that whenever economic surplus rises it's possible for those who benefit to compensate those who were harmed, and to do so in a way that ensures everyone's better off... In reality, it's rare for new policies to compensate the people they harm. Thus, the argument that it's possible to make everyone better off is just that, a possibility...consequently, real-world policy debates are rarely just about efficiency. They also focus on equity (Stevenson & Wolfers, 2019, Section 7.1).

While this logic can be helpful, it is also an attempt to have our cake and eat it too. Thinking of of an outcome as "more efficient" because it is closer to an outcomes where transfers could make everyone better off, but doesn't, stretches the term efficiency beyond any meaningful sense of the word. If we are not actually achieving a Pareto gain then we are

[6]One way past this would be to assume all utility possibility frontiers are parallel, but this requires unrealistic restrictions on preference heterogeneity (Gorman, 1955).

not avoiding interpersonal comparisons or sticking to ordinal preferences. Weighing a gain in consumer surplus against a change in economic inequality, for example, requires assuming that surplus corresponds to welfare or well-being in a way that can be compared across people. This is a fine and common assumption to make, but it is not an efficiency argument.

Hendren and Sprung-Keyser (2020) provide a helpful framework for thinking about this type trade-off. They propose measuring the marginal value of public funds for individual policies and using the following logic for policy comparisons: "Given two policies, A and B, suppose MVPFA = 2 and MVPFB = 1. Then one prefers more spending on policy A financed by less spending on policy B if and only if one prefers giving $2 to policy A beneficiaries over giving $1 to policy B beneficiaries." As I show in section A.2, this logic works well if policies impact homogeneous groups but places a high informational burden on policymakers for heterogeneous populations. Most importantly, it again assumes utility is cardinal and corresponds to individual well-being.

While economics can analyze and frame policy analysis in various ways, interpersonal comparisons are an unavoidable aspect of collective decision making in most practical applications. One way to do this is by building on the preference utility framework and assuming that individual behavior is the result of maximizing cardinal utility where utility corresponds to a meaningful measure of welfare or well-being. What does this assumption imply about how utility changes with income or across individuals? How does this assumption fit into the re-distributive preferences of policymakers? Are policymakers true utilitarians under this definition of utility, or do they consider the total distribution of utility as well? Out understanding of this assumption and it's implications would be greatly improved if we knew what it implied about the concavity of utility across income. In the next section, I state and explain the assumptions needed to measure that concavity.

3 Identifying Assumptions

The empirical model is incredibly simple. So simple, in fact, it glosses over some important assumptions that are implicitly being made. Like, for example, what is utility? The following assumptions give the theoretical requirements necessary to identify the marginal utility of income using the strategy outlined in section 5

The first assumption needed for identification is the following

Assumption 1. *The Marginal utility of the pain relief described in the survey questions is independent of income.*

So if we look across income groups, the average utility from pain relief in the survey

scenarios would always be equal. We can see this in the diagram in figure 1 represented as equality 1 (corresponding with assumption 1). The estimation strategy will work for any conception of utility that satisfies this equality, and I expect it to hold for a wide variety of conceptions. This means the results are broadly applicable and not subject to slight deviations in opinions about how interpersonal comparisons of utility should be made. That being said, what exactly does it mean for utility to be equal across people?

Utility is a concept that is frequently used in economics and philosophy but it's definition is not consistent and it is often used in differing and conflicting ways. Assumption 1 requires a cardinal definition where the experiences of two different people, or the average of two groups to be more exact, can be considered equal. However, even among cardinal definitions of utility there is a wide variety of views about what exactly utility is.

Hausman and McPherson (2006) provide a clarifying framework for grouping theories of well being. Substantive theories say *what* things are inherently good. For example happiness, or pleasure could be considered the actual meaning of utility. The substantive approach fits with the work of utilitarian philosophers Jeremy Bentham, John Stuart Mill, and Henry Sidgwick (Driver, 2014). These theories fit well with assumption 1. It is reasonable to think, for example, that two people hitting their shins are experiencing the same pain or lack of pleasure.

Formal theories instead specify *how* to find out what is good, but not what those good things actually are (Hausman & McPherson, 2006). Welfare is considered to be the satisfaction of preferences and utility is the extent to which those preferences are satisfied. Irving Fisher actually preferred terms like "wantability" to avoid conflation of what economics is considering utility with the substantive theories of "Benthom and his school"(Fisher, 1927). However, some moral philosophers have accepted this as their preferred understanding of utility (Hare, 1981). If utility is the satisfaction of preferences how can we compare that across people?

Harsanyi (1955, 1986) proposed a way to think about how to compare the intensity of preferences across people using "extended preferences". This idea was even considered by Arrow (1978) as a way past the independence of irrelevant alternatives assumption in his impossibility theorem (Arrow, 1950). The idea proposes a thought experiment for interpersonal comparisons of utility. To compare the utility cost of two people hitting their shins on a coffee table we must imagine being each person and hitting our shin and compare those experiences. Harsanyi (1986) explains further that you have to imagine yourself with that person's preferences. MacKay (1986) calls this the mental shoehorn trick (Hausman & McPherson, 2006).

While this seems to be the dominant theory for rationalizing interpersonal comparison of

preferences, it is controversial. Arrow (2012, pg 115) says of it, "The principle of extended sympathy as a basis for interpersonal comparisons seems basic to many of the welfare judgments made in ordinary practice. But it is not easy to see how to construct a theory of social choice from this principle." Moreover, theoretical work has considered it unsatisfactory (Hausman, 1995; MacKay, 1986). The advantage of the particular assumption in this paper, however, is that we need not be capable of comparing every possible situation, only the ones in my survey which have been specifically chosen for their ease of utility comparisons. These scenarios are cases that I expect fall into the basic welfare judgments arrow referred to. Imagining oneself as a rich person hitting their shin and a poor person hitting their shin seems like a reasonable task. Beyond just being a reasonably simple comparison, I expect the experience to be much the same. I expect any of the above definitions of utility would satisfy assumption 1, which makes the estimate broadly applicable.

The medical literature does provides some helpful references for substantiating the belief that this experience would be the same regardless of income. The International Association for the Study of Pain (IASP) defines pain as "An unpleasant sensory and emotional experience associated with actual or potential tissue damage, or described in terms of such damage" (Aydede, 2019). Under this definition, people have "epistemic authority with respect to their pain: they seem to be incorrigible, or even infallible, about their pains and pain reports" (Aydede, 2019). This means pain can not be measured with a medical instrument, it must be elicited from the person experiencing it. The pain sensitivity questionnaire (PSQ) is a validated measure of pain sensitivity. First, respondents were asked to fill out the survey questions. Then, their responses are compared with questions asked during actual painful experiences (heat, cold, pressure, and pinprick) (Ruscheweyh et al., 2009).

While the very nature of pain makes it difficult to measure, the medical literature seems to suggest that, if anything, lower income people are less tolerant to pain. Miljković et al. (2014) shows that participants with a lower household possession index, suggesting lower income, are more sensitive to pain. Research has also shown that chronic pain increases pain sensitivity (Ruscheweyh et al., 2012). These results align with the idea that richer folks may have access to goods that decrease pain sensitivity. While nothing can immediately relieve the pain from bumping my shin, perhaps the knowledge that I can reward myself for enduring that pain with an expensive treat makes the experience less difficult. Moreover, the research on chronic pain suggests that perhaps rich and poor folks are not coming from the same starting quantity of daily pain. If there are diminishing returns to pain relief, this would impact that benefit of relief from a marginal painful event.

If these relationships are true, my estimates will be biased since low income folks will be getting more of a benefit from pain relief. Despite the findings described, I do not find

the medical literature to be conclusive. The chronic pain research, for example, suffers from selection bias. If 10 people actually have a chronic pain condition, the people that are least tolerant to pain are the ones who will actually go and get a diagnosis. This is not really addressed in the medical studies I have read probably because it is not important for their purposes. They are describing attributes of patients with chronic pain diagnosis and this is true even if it is because of selection. Additionally, studies assessing pain tolerance, like how long you can endure holding your hand in ice water may be testing a commitment to the study or looking tough as much as differences in the epistemic experience Stephens and Robertson (2020). To assuage concerns that pain tolerance may be correlated with income, my survey includes questions from the pain sensitivity questionnaire so pain sensitivity can be controlled for.

The second assumption needed for identification is about connecting our concept of welfare from assumption 1 to people's preferences and action. Specifically,

Assumption 2. *Individual behavior is the result of maximizing utility where utility holds the same meaning as in assumption 1. This implies if a person is willing to trade two goods, they provide the same cardinal utility.*

This assumption gives us the equalities labeled with a 2 in figure 1. Since people are willing to trade \$X or \$Y dollars for pain relief on average, they must provide the same utility.

For this to hold, people need to do what gives them the most utility under the same theory of utility used in assumption 1. The assumption that people maximize their own well-being often goes along with a formal preference utilitarian definition of utility. If utility is defined by whatever it is people prefer, then it makes sense that people maximize their own utility. That being said, this assumption does not does not preclude a substantive theory, such as utility being pleasure, from also holding. If, for example "happiness is the ultimate object of preference, then it could be true both that well-being is the satisfaction of preference and that well-being is happiness" (Hausman & McPherson, 2006, Pg 119)

This assumption is a utilitarian theory of action, and it is important to clearly distinguish this from a utilitarian theory of ethics. Utilitarian ethics says the ethical action is the one that maximizes the most good for the most people. A utilitarian theory of ethics would apply to a policymaker who wants to maximize collective utility. A utilitarian theory of action requires people to act in their daily lives as if they ascribe to ethical egoism, since an individual's utility can often come into conflict with what maximizes collective utility. The utilitarian theory of action is what is required by assumption 2

This is also distinctly different than saying that people's actions *can* be described *as if* they are maximizing *some* utility. If we define utility as an ordinal ranking of preferences than

it is tautologically true that people with rational preferences maximize it. Jevons reflected insightfully on this distinction, saying "Call any motive which attracts to a certain action pleasure, and that which deters pain, and it becomes impossible to deny that all actions are prompted by pleasure and pain" (Jevons, 1879; Moscati, 2018). Here he is reflecting on the Benthomite definition of utility but the same is true of others.

As I pointed out in section 2 This is not an uncommon assumption. The MVPF framework in Finkelstein and Hendren (2020); Hendren (2020); Hendren and Sprung-Keyser (2020) is a prominent example in public finance, but the consumer welfare standard in industrial organization or really any cost benefit analysis using a willingness to pay measure is implicitly making this assumption.

Despite being commonly used in economics, this is not an uncontroversial or unchallenged assumption. The potential problems come in two flavors. The first is essentially the field of behavioral economics. People may not maximize their own well-being because they are bad at making choices. They would like to maximize their own well-being, but bounded rationality or behavioral mistakes get in their way. The second is that people actually make decisions in a more complex way. Even if well informed, people may simply not act to maximize their own well-being. Sen (1977) makes this point more clear with a distinction between sympathy and commitment, "The former corresponds to the case in which the concern for others directly affects one's own welfare. If the knowledge of torture of others makes you sick, it is a case of sympathy; if it does not make you feel personally worse off, but you think it is wrong and you are ready to do something to stop it, it is a case of commitment" (Sen, 1977, pg 326). In either case, an ordinal utility defining choice would put a higher utility index on intervening to stop torture, but only in the former does the choice indicate a higher personal welfare from intervening. Assumption 2 assumes all actions are coming from sympathy. Sen goes on to point out this assumption turns people into a kind of egoist since they only engage in actions that are in their own self interest[7].

Now, believing assumption 2 is not a requirement to learn something useful from the results. It is a common assumption and understanding the implications of that assumption is useful. Additionally, models are at their core if-then statement, the results may lead readers to update their beliefs about the "then" results or update their beliefs about the "if" assumption based on the outcome. How any given reader views this test and by extension how to interpret the results largely depends on what aspect of the world is considered fixed or immovable. Is preference utilitarianism a fixed assumption, or is a particular concavity of utility, like diminishing marginal utility of income, a fixed assumption? I discuss this more

[7]They are a *kind* of egoist only, sen points out, because we aren't dictating the cause of their actions, which may not be self interest, only that the actions happen to coincide with self interest.

in section 8.

The final equality in figure 1 just follows by the transitive property and gives us the ratio of the average marginal utility of income between the two groups. While this diagram is a simple example with two groups, the same logic applies to more groups or a continuous function across income. Before building on the assumptions here for an empirical model, the following section discusses what efforts have already been made in utility measurement.

4 Previous Efforts to Measure Utility

Attempts to measure utility have a long history in economics. One of the main problems to overcome in estimating utility is that the marginal utility for any particular good will change with changes in income in unpredictable ways because of cross derivatives in cardinal utility. As I discussed above, even if everyone has the same cardinal utility function, as people gain income, the marginal utility of goods might change quite a bit depending on how consumption of compliments and substitutes changes with income[8].

While it may seem possible to control for consumption of compliments and substitutes, the curse of dimensionality will make considering all the relevant goods impossible. Not only that, but if there is heterogeneity in preferences about those goods along the income distribution it will be impossible to disentangle differences in preferences about compliments and substitutes in utility from differences in the marginal utilities of income.

Irving Fisher recognized that these cross partials are the key complication in identifying utility and proposed overcoming this problem by looking at large consumption classes like food and housing. He proposes that "the utility derived from the consumption of each commodity group depends only upon the quantity of that commodity group that is consumed." (Fisher, 1927). He is positing that the marginal utility of food, for example, must not depend on the type of housing or entertainment a person has. This independence assumption serves the same purpose as my assumption that pain relief is independent of other goods (1). An independence assumption bypasses the need to look at compliments and substitutes because there are none. However, in Fisher's case, rich and poor folks are not going to consume the same amount of quality adjusted food or housing on average. He proposes using a different location with an alternative price vector where people consume the same amount of food as the low income group, but the same amount of housing has the high income group[9]. Using this method requires solving a difficult if not impossible index number problem.

[8]These cross partials represent compliments and substitutes in utility when utility is cardinal and is how the terms were originally conceptualized (Auspitz & Lieben, 1889; Moscati, 2018)

[9]Or vise versa. The consumer with alternative prices needs to act as a link between any two broad class of goods

Morgan (1945) actually attempts to implement fisher's method and points out several difficulties. However, both Morgan and Fisher treat the required index number problem as a problem of mathematical and statistical precision. In fact, The correct index number is inextricably linked with the question of marginal utility once we allow for heterogeneous preferences across regions. Fisher's method needs the index number to show that two families "enjoy the same or equivalent food rations (Fisher, 1927)". What is the equivalent consumption of quality adjusted food for an American in dollars and an English person in pounds? The relative enjoyment of the average English person for Shepherd's Pie compared to the relative enjoyment of the average American of a cheeseburger will differ and so the ability to purchase a given consumption basket in either country does not make the utility value of a given dollar expenditure on food equivalent to the average person in the two countries. This is a general problem with index numbers or measures of inflation under heterogeneous preferences (Samuelson, 1950). We can't equate the utility consumption of food with index numbers alone and instead need a measure of the utility of food consumption to say the two people get equivalent utility from their food. Unfortunately, this is as difficult to apprehend as the original question of marginal utility of income!

Rather than rely on aggregation to achieve a good without compliments and substitutes in utility, I use a good with it as an inherent quality. This avoids the concern about different preferences for a given food price vector. Since the pain relief I describe is also not something anyone can purchase, and so the rich and poor are considering the same marginal change, I also avoid the need for comparisons across different price indices, bypassing the index number problem entirely. The core idea however, getting a good with no cross partials, is the same as in my strategy.

For an interesting and more thorough overview of this early research see (Dimand, 2019) and for a more thorough account of the history of thought on measuring utility see (Moscati, 2018).

5 Empirical Model

With these assumptions in mind we can now consider a more explicit empirical model. This model starts with an explicit functional form for cardinal utility and so takes assumption 2 as given.

$$U(m_i, X_i, \epsilon_i) = \phi(m_i) + r(X_i, \epsilon_i) \tag{1}$$

Where *phi* is a function for the marginal utility of income m and r is relief from one of

the painful experiences in my survey. I have assumed that the marginal utility of income is constant across people, but is a function of income. There are different types of people, some more or less tolerant of pain. These types are described by the characteristics in the vector X_i. Additionally, each individual's utility from pain relief has an idiosyncratic error ϵ_i. It is crucial that $m_i \notin X_i$. That is, pain relief cannot also be a function of income. It is also crucial that the utility of pain relief and income are additive separable. This is a key part of satisfying assumption 1 since this ensures there are no non-zero cross partials. With these assumptions, we can see that the reservation price for person i for pain relief is the following

$$P_i^r = \frac{r'(X_i, \epsilon_i)}{\phi'(m_i)} \tag{2}$$

The price of money is normalized to 1 and $\phi'(m_i)$ is the marginal utility of income. $r'(X_i, \epsilon_i)$ is the marginal utility from taking the pain relief in my survey questions. While this model should be familiar to anyone who has taken an intermediate economics course, the key point here is that the cardinality and separability assumptions are meant to literally hold in this particular setting rather than just act as a simplification.

Building on equation 2, the following theorem shows when marginal utility of income can be identified with a conditional average.

Theorem 1. *If* $r_i' \perp\!\!\!\perp m_i$,

then the following holds up to a normalization α

$$\mathbb{E}[P_i^r(m)|m] = \frac{\alpha}{\mathbb{E}[\phi(m)|m]} \tag{3}$$

Note that the if statement in the theorem is equivalent to assumption 1. To prove this theorem, start by taking the conditional expectation of both sides of equation 2. The expectations here are expectations across people for a given income level m. This gives

$$
\begin{aligned}
\mathbb{E}[P^r(m_i)|m] &= \mathbb{E}[\frac{r_i'}{\phi'(m_i)}|m] \\
&= \frac{\mathbb{E}[r_i'|m]}{\mathbb{E}[\phi'(m_i)|m]} \\
&= \frac{\mathbb{E}[r_i']}{\mathbb{E}[\phi'(m_i)|m]} \\
&= \frac{\alpha}{\mathbb{E}[\phi'(m_i)|m]}
\end{aligned}
$$

After conditioning on m, $\phi'(m_i)$ is a constant and so independent of r_i' and so the expectation of the product is equal to the product of the expectation. Next, $r_i' \perp\!\!\!\perp m_i$ by assumption, and so we can remove the condition from the numerator. Finally, $\mathbb{E}[r_i']$ is a constant that we can normalize to α

This normalization means we have not identified utility in the absolute sense, but in accordance with the independence of common scale, we have identified it up to a constant normalization. The ratio of $\phi'(m_1)$ to $\phi'(m_2)$ is the same for any normalization regardless of m_1 and m_2. So we can say things like the marginal utility at m_1 is twice the marginal utility at m_2 or itentify the concavity of utility. This model does not define a specific number of utils for a particular basket of goods. A useful normalization would be to choose particular income of interest. Let the average income be $\bar{m}$ and set $\alpha = \mathbb{E}[P^r(m)|m = \bar{m}]$ so $\mathbb{E}[\phi'(\bar{m})] = 1$. Now setting welfare weights equal to the inverse of marginal utility will scale WTP to what it would be if marginal utility was unchanged, but everyone had the average income $m_i = \bar{m}$.

An important point that is implicit in this model is that the price vector is fixed. In this simplistic model where money is a single good, it is hard to see, but in appendix A.1, I present the same model using indirect utility, heterogeneity, and infinitely many goods. With that approach, it is clear that the model is conditional on a price vector. This implication of this is that a change in the relative price of goods, in particular the ratio of the average price of goods consumed by the rich and the poor, will shift the marginal utility of income function. I discuss the implications of this in more detail in section 8

If income is categorical, we can simply take the average reservation price across income bins. Alternatively, the relationship between income and the reservation price can be estimating with any parametric or non-parametric estimation technique for a conditional average. For example, a quadratic polynomial. Before showing the results or building on this model, the next section reviews the data collection, population, and summary statistics.

6 Survey Data

The data for my main analysis comes from a new survey fielded to 1747 respondents through the survey panel company Centiment[10]. Respondents are recruited through Facebook, LinkedIn, and partner networks to fill out surveys for money. While not a truly random sample, My respondents are matched to the census on age, race, gender, and region and these demographics are provided to me by Centiment.

The final full set of survey questions can be seen in appendix E. The first two sets of questions are variations of validated question on the pain sensitivity questionnaire (PSQ) Ruscheweyh et al. (2009). The first set are open ended questions asking respondents to self report their willingness to pay. The second set asks a simple "yes or no" question of the form, "would you pay $X to relieve that pain", where X is randomly selected. The next set of questions is a subset of the PSQ asking respondents to say how painful scenario's would be on a 0-10 scale. Many of these overlap with the WTP questions to get a sense of individual aversion to particular events.

I chose the PSQ questions that were reported as the most severe, but also that resolve in a relatively short amount of time. The pain resolving in a short amount of time helps satisfy the assumption that the marginal utility of relief from these experiences is uncorrelated with income because other goods do not impact the experience. To better see why these are a good fit, consider an excluded question that asks about pain from a sunburn. Sunburn is much less likely to satisfy the independence assumption because the respondents job or access to soothing medication might make a difference on the actual pain sunburn causes.

One question in the first section and one in the third ask about scenarios that typically would not be painful at all. I refer to these questions as "catch questions". These "catch questions" are used to determine who is engaging in good faith in the questions and paying attention. How exactly these where used to exclude responses is outlined in section 6.1. Additionally, in the third section of the survey there is a question asking respondents to enter 9 to ensure their full attention. Respondents who did not enter 9 for this are, at that point, removed from the survey.

The next section asks a few basic demographics. Marital status, number of children, education, employment. The final sections asks about income and financial health. The financial health questions come from the Fin-health survey[11].

[10] website: `https://www.centiment.co/`

[11] WEbsite: `https://finhealthnetwork.org/`

6.1 Protest Answers and Outliers

It is common in open ended willingness to pay questions to receive protest answers to questions, yet there is not an agreed upon way in the literature of handling them (Boyle, 2017, pg.110-111). My first step to handling protest answers is removing entire respondents from the survey based on their response to the "catch" questions I mentioned above. An economist thinking about rational behavior may want to exclude anyone who did not answer 0 to both catch questions. The situations are not painful and so rational utility maximizing respondents should not pay anything to relieve the pain and should enter 0 when asked. I do not use this strict of a cutoff for my base specification. It is possible some people perceive some pain from these situations or simply did not consider 0 to be a viable option.

Instead of removing respondents with non-zero answers, I use the following conditions. First, I look at the response to the open ended catch question asking what is the most you would pay to relieve the pain of shaking hands with someone who has a normal grip. I exclude anyone who answered higher on this than any of the other open ended questions. The thought here is that, while they may not think to enter zero, it certainly should be less than or equal to questions that are clearly painful. Second, in order to check for attention, I remove anyone who answered the same thing for every question. This includes 263 respondents who entered zero for every question. I expect this in particular might be throwing away some useful information. Some people who actually have low WTP, perhaps below a dollar, may just default to zero. Especially since the question is not clear on how the payment is processed. I drop them for the main analysis since properly accounted for this truncation behavior will make things more complicated. Finally, I drop anyone who said more than $5. This is, admittedly, a bit of an arbitrary cutoff but the goal is to allow people who defaulted to low, non-zero answers without included people who where not thinking about the question.

The last condition is based on the response to the pain sensitivity questionnaire catch question. Here, I ask again about shaking hands with someone who has a normal grip, but ask them to rank the pain from 0-10. I drop anyone who's answer to this question is greater than 3. I tried a few other more complicated conditions, like being less than the mean or less than all other PSQ Questions, but the vast majority of the cases overlapped with the three or less condition anyway. Table 1 shows how many responses were dropped for meeting each condition. 1021 responses remain after all of the drop conditions.

For the open response questions, I also top coded responses. The maximum answer, for example, was one trillion dollars. Even if this were in fact true, it will throw off any analysis using means. I top-code them rather than dropping them since top-coding uses some of the information. Whoever entered such a high amount likely has a true amount that is fairly high as well. A standard rule of thumb is to add 1.5 times the inter quartile range to the

75th percentile and treat anything above that as an outlier. As my data is fairly skewed, I use 4.5 times the IQR for each income group. The number of top codes can be seen in table 2. The distributions for each open response question after removing protest answers and top coding outliers can be seen in figure 2.

6.2 Summary Statistics

Before jumping in to the analysis of willingness to pay and marginal utility, it is important to understand the population. While Centiment matched my survey sample to the census on age, race, gender, and census region, some respondents were dropped from the sample as described above. Table 3 shows rates for the matched demographic variables in my sample compared to the census. Income was not matched explicitly, but figure 3 compares the results of my survey to the CPS family income. My survey is underrepresented in the $200,000 and above categories, but is otherwise fairly representative in terms of income. Having a representative sample makes the results of the survey externally valid. The mean WTP conditional on income should represent the actual mean in the United States. However, to gauge internal validity we also want to consider how pain sensitivity differs across income.

How does pain sensitivity, measured by the mean PSQ score from 0-10, change with income? Figure 4 shows the relationship. While the point estimate for the slope is negative, it is not statistically significant. Table 4 shows the results of the simple regression from the plot. A one hundred thousand dollar increase in income is associated with a .12 decrease in average PSQ response (on a scal from 0-10). Not only is this statistically insignificant, it is practically small relative to the within income variation. Nevertheless, I do control for the average responses to these questions in section 7.2.2.

7 Survey Analysis

7.1 Open Ended Mean Results

As I showed in theorem 1, if the marginal utility of pain relief and income are actually independent, the expectation of the reservation price conditional on income gives marginal utility[12]. That conditional mean can be calculated in a variety of ways. Figures 15 though 18 show two ways to estimate the conditional mean for each question. Each plot has a point for every observation. The size of the points indicate multiple observations with the same income and WTP. The reservation prices are normalized so that the mean WTP of people

[12]up to a normalization

with incomes between \$0-25k is 1. This makes comparisons across questions of the implied ratio of marginal utility more straightforward.

The first method for estimating the conditional average WTP is to estimate the mean for four income groups. The size of each group can be seen in table 5. The average for each bin is marked with the large diamond. Bootstrap standard errors are included for each mean[13]. The second method is the quadratic polynomial shown on each plot. In addition to each individual question, figure 19 shows all four questions grouped together. Since each question is normalized by dividing by the mean of the lowest income bin, the questions are all on the same scale of relative utility. The standard errors for this plot are bootstrapped with a cluster at the individual level. Surprisingly, these figures suggest very little relationship between WTP and income.

One concern with this simple approach is that the assumption that $r_i' \perp\!\!\!\perp m_i$, in theorem 1 might not hold. While income may not itself change the marginal utility of pain relief in these situations, there may be characteristics that do impact pain tolerance and are correlated with income. Gender and age have been suggested to affect pain tolerance and by extension the utility from pain relief (Bartley & Fillingim, 2013; Lautenbacher, Peters, Heesen, Scheel, & Kunz, 2017). These are both related to income in my sample. Figure 5 shows that men are more concentrated in the high income categories. Figure 6 and table 6 show that age increases some with income, the average goes from 42 in the lowest income group to 50 in the highest. Surprisingly, neither age nor gender has a significant relationship with PSQ. That being said, The PSQ is not a perfect measure of pain sensitivity. In particular, "the most pain imaginable", may not be constant across groups. One way to further investigate pain tolerance is to look at the unconditional relationship with WTP to see if there are differences across groups. There is not a significant relationship for gender and WTP on the open ended questions. There is, however, a significant relationships with age. Older respondents are willing to pay less on average as we can see in figures 11 and 12 despite having higher average incomes. These observations at the very least, motivate adding controls as a robustness check.

Unfortunately, we cannot simply add controls as a linear parameter into our model of conditional averages across income bins or with the quadratic polynomial. This is because our outcome is in terms of WTP. This means a control for gender, for example, would give a unique intercept in terms of WTP for men and women. But, the model implies that a gap in utility from pain relief between men and women should lead to a larger difference in WTP for higher incomes. So the difference in WTP should scale with income according to the marginal utility parameter that we need to simultaneously estimate. To more appropriately

[13]Bootstrapping is preferred here given the non-normal distribution

control for these variables, our model needs a bit more structure. The next sections outline a model with controls, prove it is identified, and provide the log-likelihood functions for MLE estimation.

7.2 Open Ended Response With Controls

To move beyond simple averages, we need to further parameterize our model. As with the means, I aggregate income groups into b income bins. I assume the the marginal utility of income function is

Assumption 3.

$$\phi_i'(m_i) = \sum_{k=1}^{b} \mathbb{1}_{ik}(m_i \in k)\phi_k'$$

This gives b marginal utility of income parameters, ϕ_k', to identify. These will be the average marginal utility of income for each income bin. A continuous function, like a quadratic polynomial, could be used as well.

Now for the utility impact of r_i, I will assume X_i and ϵ_i enter the model with linear parameters in the following form

Assumption 4.

$$r(X_i, \epsilon_i) = \beta_1 + \boldsymbol{\beta} X_i + \epsilon_i \tag{4}$$

Where

$$\epsilon_i \sim \mathcal{N}(0,\, \sigma^2) \tag{5}$$

and

$$\epsilon_i \perp\!\!\!\perp m_i \tag{6}$$

Together, these assumptions give the following equation for the reservation price of pain relief

Definition 1. *Given assumption 3 and 4 the reservation price for an individual is*

$$P_i^r = \frac{\beta_1 + \boldsymbol{\beta} X_i + \epsilon_i}{\sum_{k=1}^{B} \mathbb{1}_{ik}\phi_k'} \tag{7}$$

In matrix form, the expected price vector $\mathbb{P}^r$ for the full population can be written as

$$P^r = (\mathbb{X}_i\boldsymbol{\beta} + \epsilon_i) \oslash \mathbb{M}\phi'(8)$$

Where $\mathbb{X}$ is the matrix of X_i traits influencing pain tolerance and a constant for the intercept of the numerator. In matrix form, let $\boldsymbol{\beta}$ be $[\beta_1 \quad \boldsymbol{\beta}]$ from equation 7.[14] *Let $\mathbb{M}$ be an n by b matrix indicating what income bin each person i is in. Let $\boldsymbol{\phi'}$ be a b by 1 matrix of marginal utility parameters. $\oslash$ is element wise matrix division or Hadamard division.*

Appendix D proves that the model is identified, but the basic intuition is that it is identified so long as demographic characteristics do not show up in both the numerator and the denominator. For example, differences in pain relief across age cannot be separately identified from differences in marginal utility of income across age. This is why the independence of income and the utility from pain relief is crucial.

7.2.1 Likelihood Functions

We can estimate the model with MLE. Definition 1 gives the following probability density function

Definition 2.

$$P(P_i^r|\mathbb{X}_i, m_i \in k) = \pi \frac{1}{\sqrt{2\pi \frac{\sigma^2}{\phi'_{k_i}}}} e^{-\frac{(P_i^r - \hat{P_i^r})^2}{2(\frac{\sigma}{\phi'_{k_i}})^2}} \tag{9}$$

and the following log likelihood for a population of size n

Definition 3.

$$L(\mathbb{X}_i, m_i) = -\frac{n}{2} log(2\pi) + \sum_i^n -log(\frac{\sigma}{\phi'_{k_i}}) - \frac{(P_i^r - \frac{1+\boldsymbol{\beta}\mathbb{X}_i}{\phi'_{k_i}})}{2(\frac{\sigma}{\phi'_{k_i}})^2} \tag{10}$$

Where ϕ'_{k_i} indicates the marginal utility parameter for the corresponding bin for m_i. Recall that in assumption 4 the variance of the error term, and by extensions the variance in the utility from pain relief, is constant across income. This assumption allows the ϕ'_k terms to be identified off of mean differences in reservation prices, differences in the variance, and differences in reservation price gaps between control groups. We can see this since the marginal utility of income enters the same terms as σ and the term with $1 + \boldsymbol{\beta}\mathbb{X}_i$. The intuition here is that the underlying mean, variance and coefficients in utility from pain relief are constant across income. So, observed differences in the mean, variance, or coefficients for the reservation prices across incomes must be because of differences in the marginal utility of income.

[14] I realize the notation is confusing because I am using beta for two things. I'm not sure how to indicate it in a less confusing way.

7.2.2 MLE Results

I run this model with controls for age, gender, and mean PSQ. The results of the model for each open ended question are show in table 7. Keep in mind the estimates are of marginal utility ϕ_k and not WTP directly like in the unconditional means. Here, a higher value indicates a higher marginal utility. The "lemon juice in a cut" and "burn tongue on a hot drink" are both basically constant, while "touching a hot pot" and "bumping shin on a sharp edge" both actually show an increase in the marginal utility of income after \$25K. The "Hot pot" question increases to 1.74 in the 50-100K group but drops back down to 1.18 in the "more than 100K" group. A nice sanity check is that the coefficients on PSQ are positive and significant for every question. This indicates that people who are more sensitive to pain will pay more to avoid it.

I also run the model grouping the responses for all four open ended questions. This shows a slight increase in marginal utility of income in the \$25-100k range and then a slight drop above \$100k, but not as dramatic as typical assumptions would make it. Note the standard errors are in Red. I have not yet accounted for the within respondent and between question correlation (clustering at the individual level). This will almost certainly make them larger.

The contingent valuation literature suggests that open ended questions are more difficult to respond accurately to and have high rates of protest answers (Boyle, 2017, pg 110-111). With this in mind, I also ask a series of binary choice, yes or no, questions. The price proposed is randomized across respondents, but each respondent only answers each question one time, for one price. This is the opposite extreme to open ended questions in that it carries less information, but places very less burden on the respondents.

7.3 Binary Choice Model

While a binary choice question may be more familiar for respondents, leading to more accurate answers, assessing the results is more difficult and requires more assumptions. The underlying economic model is the same as in theorem 1, but now we have to model and estimate the reservation price rather than just observing it in the data. I follow the general strategy from Hanemann (1984), but the technique outlined in that paper is to calculate an overall average or median WTP. I update the technique to analyze the change in WTP with respect to income rather than a single collective estimate. The first step is to run a random utility logit regression of the following form.

$$V_i = \sum_{j=1}^{4} \delta_j \mathbb{1}(M_i = j) + \gamma X_i + \sum_{j=1}^{4} \beta_j \mathbb{1}(M_i = j) * P_i + \epsilon_i \tag{11}$$

Where δ_j is an intercept coefficient for income level j, γ is a vector of coefficients for controls X_i and β_j is the price coeffficient for income group j. It might seem incorrect, at first pass, to include unique intercepts for income in the utility model. This seems to imply different utility levels across incomes (the opposite of the identifying assumption). However, it is important to remember that V_i is ordinal utility that should not be compared across individuals and is, in my opinion, better thought of as just modeling choice probability. The differing utility intercepts allow for different choice probability levels across incomes while the differing price coefficient allows the choice probability to decrease deferentially with price across incomes. The intercepts allow the model to more flexibly estimate WTP with less strict functional form assumptions. A nice clarifying example is to imagine a good where people of different incomes buy it with the same probability at a price near zero (indicating a similar choice probability intercept). This observation would not in anyway indicate that the good provides the same marginal benefit to both people. Similarly the converse, having different purchase probability intercepts, does not imply different marginal benefits.

Putting our above equation in terms of Hanemann (1984), let α_1 and α_0 be the utility from pain relief and no pain relief respectively. We can write the difference in utility from paying for pain relief as

$$\Delta V = (\alpha_1 - \alpha_0) - \beta_j P \tag{12}$$

Now $Prob_1 = F_n(\Delta V) = (1+e^{-\Delta V})^{-1}$ is the probability of purchasing relief. This gives a distribution of the reservation price of $1 - F_n(\Delta V) = 1 - G_{P_r}(P)$. Now the mean reservation price, $\bar{P_r^j}$, for an income group j is represented by

$$\bar{P_r^j} = \int_0^\infty [1 - G_{P_r}(P)]dP \tag{13}$$

Connecting this back to the actual logit model we can write this as

$$\bar{P_r^j} = \int_0^\infty [1 - \frac{1}{1 + e^{\delta_j + \bar{X}\gamma + \beta_j P}}]dP \tag{14}$$

Where $\bar{X}$ is the average value of the controls, age and gender and mean PSQ in this case, for the entire population. This assigns each group the ordinal utility for the average age and gender, but varies the ordinal utility associated with the income group and then scales the ordinal utility by the income specific ordinal marginal utility of income. The difference across groups shows the difference in WTP attributable to changes in income.

A slight variation I also show in the results, which is used in Bishop and Heberlein (1979) and mentioned in Hanemann (1984), is to cap the integral at the maximum price. I report both values.

7.3.1 Binary choice Results

Figures 20 to 22 show the percentage of respondents that accepted the price at each income level and at the various prices. We can see that, generally, the probability of accepting the offer decreases with the price, but a relationship with income is not clear. Table 9 shows the results of the binary choice model described above with controls for age, gender, and the mean response to the 0-10 PSQ questions. Confidence intervals are calculated by taking the corresponding percentile in a bootstrapped sample of estimates. The mean and truncated mean are what are described above, but I also report the median estimate described in Hanemann (1984). The first two questions seem to support the same story as the open ended responses while the last suggest an increased WTP, and so decrease in marginal utility. Between the lowest and highest income groups the WTP doubles if we look at the means and increases five times for the median. The confidence intervals for all of these estimates, however, are quite large. This is the biggest limitation of the binary choice approach.

8 Discussion of Results

While there is some indication of diminishing marginal utility in one of the binary choice questions, the broad implication of the results is that marginal utility of income is constant across income groups, or at least diminishes slower than expected, implying that cardinal utility is roughly linear in dollars. This is a surprising outcome and certainly does not match my hypothesis going into this project.

To better conceptualize the implications for this result, it's important to remember the core assumption it rests on. Assumption 2 says preferences corresponding to a cardinal utility representing well-being. So to be completely accurate, what I have shown is that cardinal decision utility corresponding to individual well-being is not consistent with diminishing marginal utility of income. As I see it, there are two potentially reasonable responses to this information. One is to continue assuming that preference utility is a clear and accurate description of well-being and so utility must be linear. The other option is to say that any welfare relevant definition of utility would have diminishing marginal utility of income. So given that utility under assumption 2 does not have this property, than this must not be a welfare relevant definition of utility[15]. People may be rational, but this would mean the results are evidence that they do not maximize their own well-being. This is a more radical

[15]These are generally speaking the options for how to respond to any model. Models that rely on assumptions are, at their core, if-then statements. Here I am saying we can either accept the if and the then conclusion, or consider the contrapositive. Since we did not get the "then" statement we believe to be true, the if condition mush not hold.

departure from current practice in normative welfare economics, but it certainly has some support even within economics. I discuss each option and their implications in turn.

First, what it would mean if marginal utility of income was not diminishing? The following welfare function will help to distinguish between two possible interpretations of a constant marginal utility of income.

$$\sum_i W(m_i) = \sum_i \gamma(U(m_i)) \tag{15}$$

In this equation, $W(m_i)$ is the total weight the social planner places on a person with income m_i. $U(m_i)$ is the utility at income level m_i and γ is a function that expresses the policymakers preferences over the distribution of utility. So, A pure utilitarian would have $\gamma(U(m_i)) = U(m_i)$ and value only the sum total of utility. Other welfarists might have a preference over not just the sum total utility, but how that utility is distributed. Suppose even though I know it will hurt just as much to pump their shin, I would rather a richer person do it than a poorer person since the richer person is at a higher utility level. In this case $\gamma(U)$ will be concave in utility.

The results in this paper provide information about the shape of the $U(m_i)$ function, implying $U(m_i) = m_i$. What this implies about policymakers depends on how much we know about W compared to γ. Suppose we know a lot about policy-maker's preferences over the distribution of utility, but not much about their preferences over the end distribution of money. In other words, $\gamma(U)$ is fixed and known. In this case, the knowledge that $U(m_i) = m_i$ will update our beliefs about policymakers total welfare function $W(m_i)$. A pure utilitarian, for example, has $\sum_i W(m_i) = \sum_i U(m_i) = \sum_i m_i$. For an egalitarian the impact on the concavity of W is actually ambiguous. Consider a welfarist with a strong equity preference. Similarly to the utilitarian, a linear utility function makes the egalitarian realize that marginal dollars to the wealthy create more utility than they thought. However, the impact on the *distribution* of *total* utility is unclear. If the slope of the linear utility is large relative to their priors, total utility may be much more unequal than they thought and push them towards more redistribution. If the slope is small, the world is less unequal and would push them towards less redistribution. How exactly one should think about the level changes in utility across income is still completely subjective.

While I found this line of reasoning to be the more common first reaction, on further inspection, it might not be the best choice. I am not aware of work that elicits views about preferences over the distribution of utility asks if policymakers really subscribe to a welfarist or utilitarian ideology in accordance with assumption 2. Economic work has asked how much redistribution people would like and has inferred it from existing policy (Hendren,

2020; Kimball, Ohtake, Reck, Tsutsui, & Zhang, 2015). What this shows is that most people think giving $1 to a poor person is better than giving it to a doctor. People do use terms like "it would make a bigger difference to the poor person", but it's not actually obvious what they mean by that. Do they mean the poor person desires it more in the utilitarian sense, or do they mean it in some other way? Perhaps they see the things its being spent on as more worthy or more important beyond simple desire. In this case, it makes sense to think of their preference for redistribution, W, as fixed and update our beliefs about $\gamma(U)$. This shows us how their motivations fit into the welfarist framework build on economic assumptions. This approach indicates that the $\gamma(U)$ is more concave than typically assumed. In other words, the welfarist, but non-utilitarian, portion of people's re-distributive preferences is larger than expected. Very few people subscribe to a pure utilitarian view under the economic definition of utility since very few people favor no redistribution.

Is a constant marginal utility of income actually plausible? Is this a reasonable assumption to drop? The idea bumps up against some common intuition but may not cause as much conflict as one would expect. Constant marginal utility of income does not mean that there is not still diminishing marginal utility for particular goods. The tenth apple is likely less enjoyable than the first. Part of the reason for this is the assumption that people have convex preferences and so would prefer to spread their money out on other goods rather than concentrate it on apples. This idea does not apply to money since doubling money would allow a consumer to double everything in an optimal way in accordance with convex preferences. On the other hand, the logic for the diminishing marginal utility of single goods may also come from the idea that goods generally become less desirable, in a well-being sense, the more a person has. In this case, doubling my income and doubling everything I own would lead to less than double utility. In the real world, if someone's income were to double, they almost certainly would not just double the same goods they already purchase. People don't have homothetic preferences and so the quality and type of goods might change significantly as their income changes. A linear marginal utility of income actually has no bearing on the marginal utility of any other given good. In fact, if we see that people are willing to pay less in dollars for additional goods, we can now confirm that those goods do in fact have diminishing marginal utility.

Non-homothetic preferences play another key role in why linear utility is plausible. The average marginal utility of income depends on the price vector in the economy. Suppose necessities become more expensive and the consumer goods available at low incomes shift to a lower quality. Suppose at the same time efficiency gains are coming to expensive luxury items and services that the low income workers performing those services could never afford themselves. In this case a dollar might just buy a lot more value in the hands of a wealthy

person. This idea is also exacerbated by the related idea that it is expensive to be poor. Berkouwer and Dean, for example, show that households in Nairobi are only willing to pay $12 for a stove that would save $237 over two years, and that a low interest loan increases willingness to pay to the actual savings over the life of the loan (Berkouwer & Dean, 2021). Lower income folks may end up buying lower quality goods that don't last as long and so receive less consumption per dollar.

Another important consideration for the plausibility is how to reconcile a linear utility function with risk aversion. A constant marginal utility of income would imply risk neutrality. Here I think behavioral economics can easily square the circle. For example, reference dependent utility can make a big difference. People might assume that if they won $100,000,000 they would suddenly be much more flippant with money because it no longer matters. This expectation would influence their risk taking behavior even if, in the even they win, they update their reference point and still desire $100 about the same. This may also explain how the common intuition about diminishing marginal utility of income could be so off.

Our intuition about what utility is under assumption 2 may just not be very good. Under this assumption, the marginal utility of income is really about people's attitudes towards spending. How willing is a given consumer to part with money for a given gain? How flippant are they with cash? How much do they desire money relative to particular goods? This desire just might not stack up well with policymaker's ethical beliefs about who is deserving of this money. As explained above, we can adjust the $\gamma(U)$ in our welfare functions to match policymaker perceptions of fairness to the implications of assumption 2. This is possible so long as individuals are still maximizing their well-being. but, what if desire or a persons actions are not even internally consistent with what is good or welfare maximizing? This cannot be reconciled by manipulating $\gamma(U)$. This irreconcilability is the basis for the alternative interpretation of the results. Evidence that preferences do not reveal cardinal well-being.

Sen (1977) presents the issue with assuming preferences reveal cardinal well-being quite clearly

> A person is given one preference ordering, and as and when the need arises this is supposed to reflect his interests, represent his welfare, summarize his idea of what should be done, and describe his actual choices and behavior (Sen, 1977).

A single utility function to describe what is, ultimately, multiple different things may just be too simple. So if we actually need multiple functions, what would they all be? Behavioral economics draws a distinction between what would maximize individual well-being and what

people actual do. The underlying assumption is typically that people are ultimately making mistakes. They would like to maximize their own well-being, they just are not very good at it. There are many ad hoc models to explain errors people seem to make in particular situations, but perhaps people are generally worse shoppers than we thought? perhaps coming up with reservation prices is a difficult process and the true underlying mechanism does not relate to well-being as neatly as standard models would suggest.

Rather than it simply being too hard, people may also deviate consciously from what maximizes their own well-being. Sen (1977) draws this distinction with the ideas of sympathy and commitment. I find the following example helpful. Suppose we see a person, call them Bob, sacrifice their life to save a stranger. Sympathy would say Bob only did this because it made him better off. Maybe Bob would experience extreme guilt had he not intervened, so much so that intervening and dying was better for him. Commitment instead posits that Bob would have personally been better off had he stood aside, and yet he still intervened through an act of commitment. Perhaps a commitment to helping other whenever possible. These examples can also be practically difficult to differentiate from behavioral mistakes. Perhaps Bob thought he would live or simply did not have time to think his actions through.

One example of how actions and well-being could be systematically different would be if people have a commitment to fair pricing. Suppose people will only buy a good if it both provides more well-being than it's price *and* if the price is seen as fair. Kahneman, Knetsch, and Thaler (1991) document what people consider fair pricing to be and the results do not clearly relate to individual well-being. If the fair price limit binds prior to the welfare limit the practice of deriving welfare from the maximum willingness to pay will be hindered. This could be viewed as either a commitment to fair prices, or a behavioral mistake that people are making. Either way people are not doing what is in their best interest. I personally would have considered this to be unlikely prior to writing this paper, but it is a cogent explanation for the surprising outcome. The rich people in my survey might actually be better off if they paid more for relief, but a commitment not to pay what they view as an unfair price might hold them back.

A final and quite simple reason for why well-being and actions could differ is simple that people maximize something different than Policymakers view as valuable. As a simple case, suppose policymakers want to maximize happiness, but individuals maximize pleasure. $100 may bring more happiness to a poor person but not more pleasure which drives a wedge between preference based utility and the goal of policymakers. Replace pleasure and happiness with any two different conceptions of utility and the same problem arises.

What would accepting the idea of many different utility functions mean for economic analysis? While it might seem like serious deviation, it leaves a lot of work completely

untouched. Of course any ordinal positive observations are unchanged. Optimal tax models would need to be re-framed, but not really reworked. The social welfare welfare functions used would need to reflect the true underlying value and not necessarily revealed preferences. Policy analysis that uses a welfare framework, but not revealed preferences, would also be okay. For example, in Eastmond, Mather, Ricks, and Betts (2022) we use a welfare framework weighting changes to test scores according to their impact on student well-being. The students don't need to desire higher test scores and so nothing is changed. The area that would be significantly disrupted is normative analysis like Finkelstein and Hendren (2020), where willingness to pay is summed to measure the welfare impact of a given policy since people may be making decisions based on commitment or behavioral mistakes, for example. There may very well be ethical reasons to give people what they want even if it doesn't maximize well-being, but it would require a different ethical approach than the welfarist framework typically employed.

A third possibility that I have not yet touched on is that the results are just biased. I certainly do not see this as the most likely outcome, but I do want to touch on the weakness so that future work can improve upon the methodology.

The biggest area of concern is selection bias into the survey. What kind of rich person does a survey for a few dollars? One who is willing to do things for little money. This is a characteristic that will lead them to systematically give lower answers than their average peer. This motivated me to originally go with the company pollfish. They collect responses from phone app users and pay the users with "in app" benefits related to the apps purpose like a free yoga lesson or news article. My expectation is that a wealthy person is more likely to do a survey to get free lives while playing a game like candy crush, viewed more as an ad experience, than to sign up to do surveys for money, viewed more as a job. This is speculation on my part, and the sample in the pilot I did with Pollfish was not representative on observables of the United States. The non-Representative sample is what convinced me to switch to Centiment. However, the results from the pollfish survey show an increase marginal utility of income with income. Appendix F shows the mean figures ad MLE results for the questions in my pilot. For the pain questions, the marginal utility for the 0-25 group was high, which I expected was due to non-working spouses or temporary unemployment making up a disproportionate share of that group. However, after this initial income group, we see an increase in WTP, implying a decrease in marginal utility of income. The sample is smaller and does not reflect national demographics, but the difference across these two panels suggests more work is warranted. Another way to address this issue in the future would be to increase the reward for participation in the survey.

Another clear concern is that it's an elective and hypothetical survey and not properly

incentives. This is partially alleviated because the analysis does not require the absolute value of reservation prices to be at all accurate. It just needs to be the case that the ratio across incomes is consistent with real purchasing activity. This issue could be solved with an incentivized experiment, which I hope to follow up with inn the future.

It is also possible pain tolerance changes significantly across income. Beyond showing that the PSQ does not change across income, there is not much that is possible to say on the matter. The best way to address this concern would be with follow up studies using other goods that are conceivably uncorrelated with income. in the Pollfish pilot I also asked question about disgusting scenarios. The variance in these questions was much higher. My intuition is that there is much more variance in what people find disgusting and how much of it they can handle. This convinced me to focus on the pain questions in my final survey in order to make the most of my limited budget, but perhaps another look at things like disgust is warranted in future work.

9 Conclusion

Measuring utility is a difficult problem with a long history. Without any structure, measuring the marginal utility of income is not even a well defined problem , but with the right structure and example it is empirically possible. I present a model for a feasible identification strategy for the average marginal utility of income and implement that strategy with a new survey. The results suggest that the marginal utility of income is constant across income groups, implying that utility, which corresponds to preferences and well-being in a cardinal sense, is roughly linear in dollars. The surprising result can be also be framed as a dilemma between accepting linear utility, or revoking our confidence in the the common assumption that preferences correspond to well-being. Either option is a logically consistent way to justify the findings, and either option has significant implications for our understanding of individual preferences, utility, well-being, and distributional ethics.

Appendix A Theory Extensions

A.1 Generalizing the empirical model

The simple empirical model laid out in the body of the paper includes money as a numeraire good. However, the same estimation strategy can be supported using heterogeneous preferences and indirect utility. I think this model is more accurate to the real world where there are many goods and a whole set of prices, but the estimation is ultimately identical and it is more confusing. Thus, it is here in the appendix.

First consider an indirect utility function V that is a function of prices P and income y, but also other characteristics θ like gender, age, pain sensitivity, or anything else that might influence preferences. This allows V to be heterogeneous across different people. This gives

$$V(P, Y, \theta) \tag{16}$$

Now for the pain relief described in our questions needs to be re-characterized as a price change. Let P be the price vector in our current world, and let P' be a price vector where immediate pain relief is free. Now we can define the change in utility from pain releif as follows.

$$V(P', Y, \theta) = V(P, Y, \theta) = \Delta V^j \tag{17}$$

Since these are relatively small changes we can, as in the empirical model in the body, treat this as a marginal change. That means the following equality holds

$$EV(P', P, Y, \theta) U'_y(P, Y) = \Delta V^j \tag{18}$$

Where U'_y is the marginal utility of income and EV is equivalent variation. To shorten the notation, let $EV(P', P, Y, \theta) = EV^j$.

An important note regarding the previous step is that the marginal utility of income is not a function θ. So, while the preference for pain relief may vary by characteristics like age, the marginal utility of income cannot.

Now to identify the marginal utility of income we need the same assumption as in the simpler model. That is, we need the utility from pain releif to be indpenet of income. With that we get the following

Theorem 2. *if* $\Delta V^j \perp\!\!\!\perp Y$ *then*

$$\mathbb{E}[U'_y(P, Y)|Y, P] = \frac{\alpha}{\mathbb{E}[EV(P', P, Y, \theta)|Y, P]} \tag{19}$$

Proof. Given $\Delta V^j \perp\!\!\!\perp Y$ we get that

$$\mathbb{E}[\Delta V^j|P,Y] = \mathbb{E}[\Delta V^j|P] = \alpha \tag{20}$$

□

The first equation follows from the independence of the change in utility, ΔV^j and income Y. The second equality is a bit of an abuse of notation, but here the price vector is actually set to our current price vector P, and so this is a constant we can normalize to any level. Now, given this equality we get

$$\begin{aligned}
&\alpha && (21)\\
&= \mathbb{E}[\Delta V^j|Y,P] && (22)\\
&= \mathbb{E}[EV(P',Y,P,\theta)U'_y(Y,P)|Y,P] && (23)\\
&= \mathbb{E}[EV(P',Y,P,\theta)|Y,P]E[U'_y(Y,P)|Y,P] && (24)
\end{aligned}$$

The second line is the normalization explained above, the third line is the identify also explained above, the last line comes from the fact that $U'_y(Y,P)$ is only a function of Y and P and so conditioning on those makes it a constant. Rearanging the equation gives us the theroem.

One point this more complex model makes clearer is that the model estimates the marginal utility of nominal dollars, conditional on a price vector. A nominal dollar might not provide the same purchasing power to everyone in a world of non-linear pricing, quality variation, credit constraints, non-convex preferences, and other complications outside of basic economics models. In particular, a marginal dollar may have more, or less, purchasing power the more dollars someone has. A simple example would be to consider geographic sorting. Suppose richer people live in more expensive areas. In this case richer people value dollars less because the marginal value of consumption is lower, but also because a dollar literally buys less at stores in their area. This example might theoretically be controlled for with geographic price indices, but other examples are more complicated.

The following story popularized by novelist Terry Pratchett illustrating why the rich are able to spend less. Suppose a quality pair of boots that will last ten years is $50, but a cheap pair that will last only a year is $10. A poor person, with a marginal dollar, may only be able to purchase the cheaper option despite it being more costly in the long run (Flood, n.d.). The poor person may appreciate the boots more, in line with diminishing marginal utility of consumption, but the richer person is able to purchase boot years at half the price,

making their marginal consumption per dollar higher. This example supports the colloquial saying[16], "it's expensive to be poor". Higher credit rates, an inability to buy in bulk or take advantage of off peak sales might mean the poor just can't buy as much with an additional dollar. Berkouwer and Dean, for example, show that households in Nairobi are only willing to pay $12 for a stove that would save $237 over two years, and that a low interest loan increases willingness to pay to the actual savings over the life of the loan (Berkouwer & Dean, 2021).

Perhaps a more fundamental consideration is that consumption is not homothetic and so the types of goods people consume with their first $10,000 a year look very different than after making $100,000. The first priorities for consumption are essentials like food and shelter. If food staples and housing become more expensive relative to luxury or entertainment goods, than we will see the marginal utility of a dollar for the poor fall relative to the rich. So, in general, the marginal utility of a nominal dollar also captures the relative cost of the differing consumption baskets of each income group. Interestingly, looking at these utility estimates over time could capture to what extent inflation has been concentrated on essentials or low quality items compared to luxury goods.

A.2 Using Condition Expected Marginal Utility

One goal of the empirical methodology I develop in his paper is to better understand and use welfare weights for real world policy analysis. The model below shows the general case for a policy change with completely unrestricted heterogeneous individual preference. That is, preferences can be so heterogeneous that literally every person is different. In the case, which is almost surely the case in real life, is there any hope to actually recovering the welfare impact of a policy change even if we know the conditional marginal utility of income? I expand on the approach from Hendren and Sprung-Keyser (2020) to show show how estimating a conditional WTP lowers the informational burden on policymakers. This compliments the empirical exercises in this paper which is estimating the conditional marginal utility of income.

To start I want to show the limitations of using the sum of WTP to estimate welfare. Suppose cardinal indirect utility is represented by

$$V(P, Y, \theta) \tag{25}$$

where Y is a vector of "observed by the econometrician" terms that impact marginal utility of income (income, financial health, race, in my case), θ is a vector of "un-observed

[16] I've at least heard this a lot among family and friends. Not sure how common it actually is

by the econometrician" terms that influence preferences (and may also influence the marginal utility of income). This can be things like color preference, but also can be as idiosyncratic as individual indicators for an entire individual specific ordinal preference function over all goods. The later example demonstrates that heterogeneity in individual preferences over goods is not restricted in any way with this general parameterization.

Now a social planner is considering a policy j that will change prices from P^0 to P'. The social planner wants to know the sum of the welfare weighted change in utility from their proposed policy j,

$$\sum_i \gamma_i \big(V(P', Y_i, \theta_i) - V(P^0, Y_i, \theta_i\big) = \sum_i \gamma_i \Delta V_i^j \tag{26}$$

where γ_i is the welfare weight they place on individual i's utility.

Now let $MUy(P, Y, \theta)$ be marginal utility of income s.t.

$$EV(P^0, P', Y_i, \theta_i) MUy(P^0, Y_i, \theta_i) = \Delta V_i$$

where EV is the equivalent variation for policy j. Let EV_i^j be shorthand for $EV(P^0, P', Y_i, \theta_i)$. Now how does WTP or the conditional marginal utility of income fit into this? If we expect policy makers to make use of simpel total WTP estimates, we are using the approach outlined in Hendren and Sprung-Keyser (2020) which gives the following equation.

$$\sum_i \gamma_i \Delta V_i^j = \sum_i \gamma_i MUy(P^0, Y_i, \theta_i) EV_i^j \tag{27}$$

$$= \frac{1}{\sum_i EV_i^j} \sum_i \gamma_i MUy(P^0, Y_i, \theta_i) EV_i^j \sum_i EV_i^j \tag{28}$$

$$= \bar{\omega}^j \sum_i EV_i^j \tag{29}$$

This reorganizing of terms makes it seem like $\sum_i EV_i^j$ will be an extremely useful metric for policymakers since all they need to know is $\bar{\omega}^j$ which is partially normative since it contains the welfare weights γ_i. Even if the policymaker knows γ_i and $MUy(P^0, Y_i, \theta_i)$, it's not clear that they would know $\bar{\omega}^j$ since it includes the interaction of each individuals marginal utility and their equivalent variation for the policy. So knowing $\bar{\omega}^j$ requires knowing about the joint distribution of the policy impact and the marginal utility of the people impacted. This is a high informational burden!

There is a special case where the informational burden is low. Consider the following

Theorem 3. *If $\gamma_i * MUy_i^j$ and EV_i^j are independent then*

$$\sum_i \mathbb{E}[\gamma_i \Delta V_i^j] = \sum_i \mathbb{E}[\gamma_i MU y_i^j EV_i^j]$$
$$= n * \mathbb{E}[\gamma_i MUy(P^0, Y_i, \theta_i)]\mathbb{E}[EV(P^0, P', Y_i, \theta_i)]$$

Cases where this hold would be policies that impact people essentially at random. Another case is where a policy impacts a very homogeneous subgroup that all have similar $\gamma_i \Delta V_i^j$. A welfare policy targeted at single parents with a specific income level is a good example that approximates this situation. Since the welfare weighted marginal utility is essentially constant, than it is independent of the equivalent variation of the policy. In these cases we can multiply the sum of WTP by the average welfare weight for the group, which is the process Hendren and Sprung-Keyser (2020) suggest, and recover welfare.

While this sometimes works, knowing the conditional marginal utility of income allows us to relax this assumption and recover welfare in more cases. With conditional marginal utility, we get the following

Theorem 4. *If $\gamma_i * MUy_i^j \perp\!\!\!\perp EV_i^j | P, Y$*

$$\sum_i \mathbb{E}[\gamma_i \Delta V_i^j] = \sum_i \mathbb{E}[\gamma_i MU y_i^j EV_i^j | P, Y]$$
$$= n * \mathbb{E}[\gamma_i MUy(P^0, Y_i, \theta_i)|P, Y]\mathbb{E}[EV(P^0, P', Y_i, \theta_i)|P, Y]$$

Now we can still recover welfare from policies that have differential impacts across Y so long as there is no joint distribution between the marginal utility and the EV of the policy in the θ terms. For example, the policy can have differential impacts by income, observed and conditioned on in y, but not color preference, unobserved and unconditioned in term in θ. An example where this would not hold, and so separating the expectations biases the estimate, is if a policy disproportionately helps people who like the color red and we live in a world where red things are expensive and so people who like red have lower marginal utility of income. Color preferences is intentionally pedantic because it demonstrates the difficulty of incorporating heterogeneous preference into a welfare model. We want to allow people to have different color preferences. However, if we don't want to restrict preferences, this also means that if the price of red things increases, the marginal utility of income for people who

like red decreases faster than their peers who do not like red. This makes recovering welfare in general seem impossible since so many factors influence the marginal utility of income. On the contrary, this is in fact a mathematical distraction. If red dye is more expensive than blue, should the government transfer money from people who like blue to those who like red? I've never heard of such a request. The vast majority of people simply do not see things like color preference as a valid reason for redistribution.

We can adapt the model to better reflect the actual normative preferences of most people by using a specific set of adaptive welfare weights. For simplicity, lets suppose our policymaker is a special kind of utilitarian. They want to maximize utility, but only by considering differences in marginal utility along dimensions in y, things like income wealth etc, but not along dimensions in θ, things like color preference. Then their welfare weights are the following

$$\lambda_i = \frac{\frac{1}{n}\sum_k MUy(P, Y, \theta_k)}{MUy(p, Y_i, \theta_i)} \tag{30}$$

Note that the numerator is an average of a function. What we are doing here is getting the mean marginal utility of income as a function of P and Y across all θ terms. This gives the following welfare weighted marginal utility of income for an individual

$$\lambda_i MUy(P, Y) = \frac{1}{n}\sum_k MUy(P, Y, \theta_k) \tag{31}$$

In words, this is just the average marginal utility of income across θ. Rather than try to discern the marginal utility difference between people who prefer red and blue, we just assign everyone the average. We can see how this is useful because it makes the welfare weighted marginal utility of income a function of P and Y only. Returning to analyzing the welfare of policy j, since $\lambda_i MUy$ is only a function of P and Y we get

$$EV(P, P', Y, \theta) \perp\!\!\!\perp \lambda_i MUy(P, Y) \mid P^0, Y$$

which gives

$$\sum_i \mathbb{E}[EV(P, P', Y, \theta)|P^0, Y] \quad \mathbb{E}[\lambda_i MUy(P, Y)|P^0, Y] =$$
$$\sum_i \mathbb{E}[\gamma_i \Delta V_{ij}|P^0, Y]$$

For a policymaker with these presences, this relationship holds for all polices. Now Y may contain more things than we can observe in a given case. For example it may include income

and wealth. but we only observe income in a particular data set. In this case, the same logic applies as before the welfare weight normalization. If the policy does not deferentially impact wealthy or un-wealthy people, conditioning on income will give an unbiased estimate of the welfare impact of the policy. The same idea can actually be applied to any outcome, not just WTP, to understand how to recover welfare with heterogeneity. We show this using test scores in Mather, Eastmond, Ricks, and Betts (2022)

Appendix B Tables

Drop Conditions				
All Idendical	Not the Minimum	WTP greater than 5	PSQ Greater Than 3	N
Pass	Pass	Pass	Pass	1021
Fail	Pass	Pass	Pass	258
Fail	Pass	Fail	Pass	6
Fail	Pass	Pass	Fail	22
Fail	Pass	Fail	Fail	20
Pass	Fail	Pass	Pass	19
Pass	Fail	Pass	Fail	11
Pass	Fail	Fail	Pass	47
Pass	Fail	Fail	Fail	112
Pass	Pass	Fail	Pass	84
Pass	Pass	Fail	Fail	63
Pass	Pass	Pass	Fail	84

Table 1: Each Row indicates what conditions are passed or failed. One failure leads to the response being dropped. The total for each combination of conditions are in the N column

Top Code Counts		
Question	Outliers Top Coded	Percent Top Coded
q1	62	6%
q2	58	6%
q4	56	5%
q5	74	7%

Table 2

Age Percent Comparison		
Age	Census Perc	Percent
18 to 24 years	12.0	15.8
25 to 29 years	9.2	7.7
30 to 34 years	8.8	9.0
35 to 39 years	8.4	8.8
40 to 44 years	7.9	7.4
45 to 49 years	8.1	8.8
50 to 54 years	8.3	7.3
55 to 59 years	8.6	7.1
60 to 64 years	8.0	5.1
65 to 69 years	6.8	7.1
70 to 74 years	5.2	7.3
75 to 79 years	3.6	4.5
80 to 84 years	2.4	3.1
85 years and over	2.6	0.8

Gender Percent Comparison		
Gender	Census Perc	Percent
female	51.3	49.8
male	48.7	50.2

Race Percent Comparison		
Race	Census Perc	Percent
White	75.1	74.8
Black	14.2	11.5
Native Or Pacific Islander	2.2	1.5
Asian	6.8	3.0
Other	7.4	7.1
No Response	NA	2.2

Region Percent Comparison		
Region	Census Perc	Percent
Midwest	24.0	21
Northeast	17.6	17
South	38.5	38
West	19.9	24

Table 3

Mean PSQ by Income		
Term	Estiamte	Std Error
Intercept	5.5014	0.0759
Income Thousands	−0.0012	0.0008

Table 4: Independent variable of regression is individual's average response to all 0-10 pain questions, dependent variable is income. Sample is after removing protest answers

Aggregated Income Counts	
Income	Count
0-25	242
25-50	239
50-100	334
more than 100	206

Table 5

Mean Age by Income	
Income (Thousands)	Mean Age
0-25	42
25-50	46
50-100	48
more than 100	50

Table 6

	Lemon Juice		Hot Pot		Burn Tongue		Bump Shin	
label	est	se	est	se	est	se	est	se
intercept	0.61	0.21	0.64	0.20	0.051	0.22	-0.10	0.19
sigma	1.5	0.061	1.4	0.060	1.5	0.065	1.4	0.057
Mean psq	0.13	0.027	0.091	0.027	0.20	0.029	0.25	0.027
Age	-0.0090	0.0025	-0.00059	0.0024	-0.0041	0.0026	-0.0069	0.0023
Male	0.084	0.094	-0.27	0.092	0.11	0.099	0.040	0.088
0-25	**1.0**	NA	**1.0**	NA	**1.0**	NA	**1.0**	NA
25-50	**1.1**	0.061	**1.1**	0.064	**1.0**	0.058	**1.2**	0.068
50-100	**1.0**	0.054	**1.2**	0.065	**0.99**	0.054	**1.1**	0.057
more than 100	**1.1**	0.064	**1.1**	0.065	**0.99**	0.060	**1.1**	0.064

Table 7

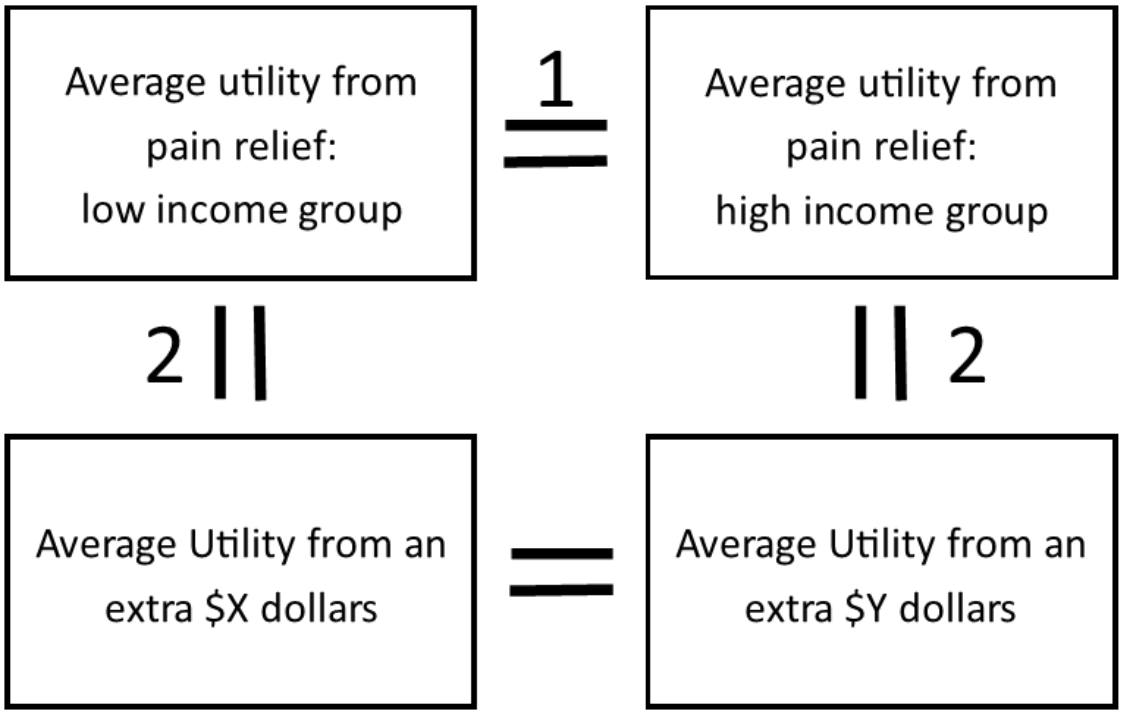
MLE Normed Pain Qs			
label	est	se	P_ratio
intercept	0.36	0.11	NA
sigma	1.50	0.03	NA
Mean psq	0.17	0.01	NA
Age	−0.01	0.00	NA
Male	0.00	0.05	NA
0-25	**1.00**	NA	**1.00**
25-50	**1.20**	0.03	**0.83**
50-100	**1.20**	0.03	**0.83**
more than 100	**1.10**	0.03	**0.91**
all questions normed to lowest mean = 1 mean psq control			

Table 8

Binary Choice With Bootstrap CI									
	Bump Elbow			Bite Cheek			Slam Finger		
	Est	LB	UB	Est	LB	UB	Est	LB	UB
meanWTP									
0-25	**25**	18	31	**36**	28	44	**69**	45	107
25-50	**37**	19	68	**41**	19	66	**115**	66	138
50-100	**25**	15	34	**33**	16	47	**131**	91	183
more than 100	**34**	19	48	**25**	19	38	**147**	123	249
trunc.meanWTP									
0-25	**24**	18	29	**35**	28	43	**68**	45	103
25-50	**31**	18	41	**41**	19	59	**111**	66	129
50-100	**23**	15	30	**33**	16	45	**120**	88	145
more than 100	**30**	19	38	**25**	19	38	**132**	116	177
medianWTP									
0-25	**10**	3	15	**16**	5	25	**39**	16	59
25-50	**8**	1	21	**19**	10	30	**76**	50	94
50-100	**3**	−2	12	**20**	12	25	**70**	42	101
more than 100	**14**	4	24	**16**	10	25	**83**	69	143

Table 9

Appendix C Figures



- X and Y are the average reservation prices for each group respectively

Figure 1

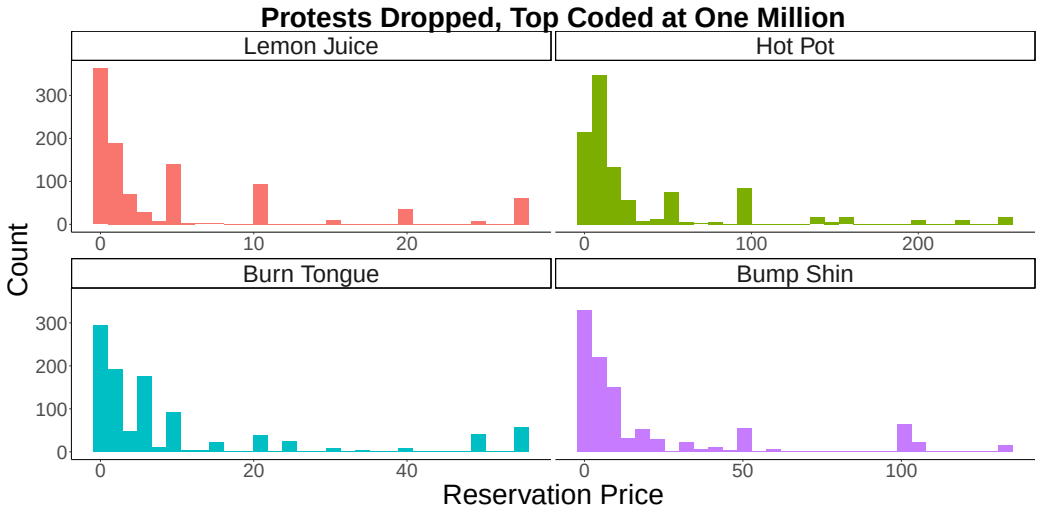


Figure 2

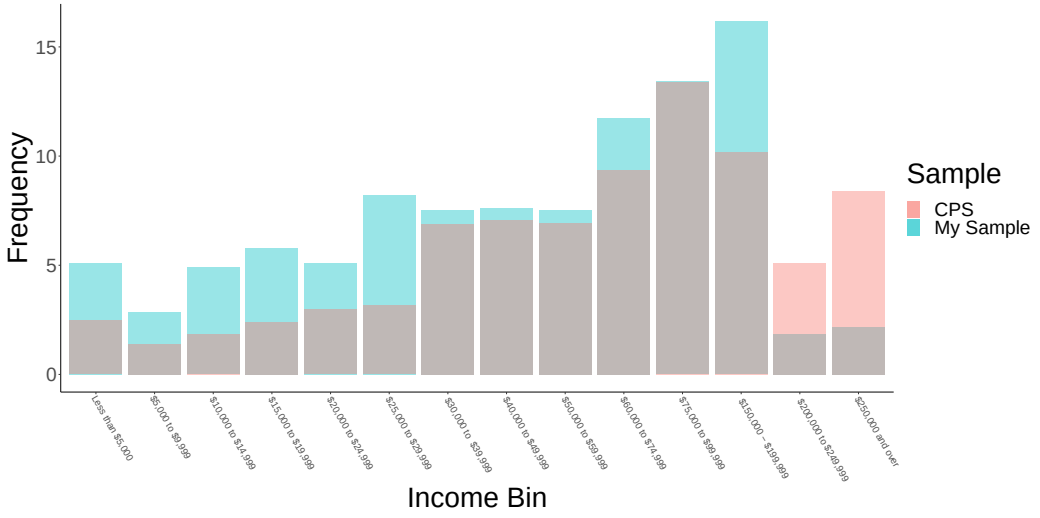


Figure 3

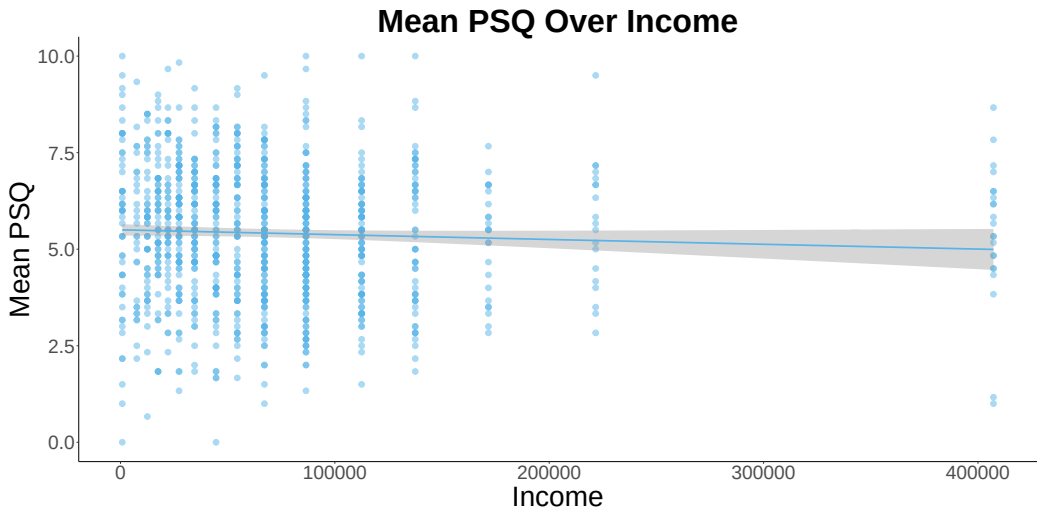


Figure 4: each individual's average response to all 0-10 pain questions after removing protest answers

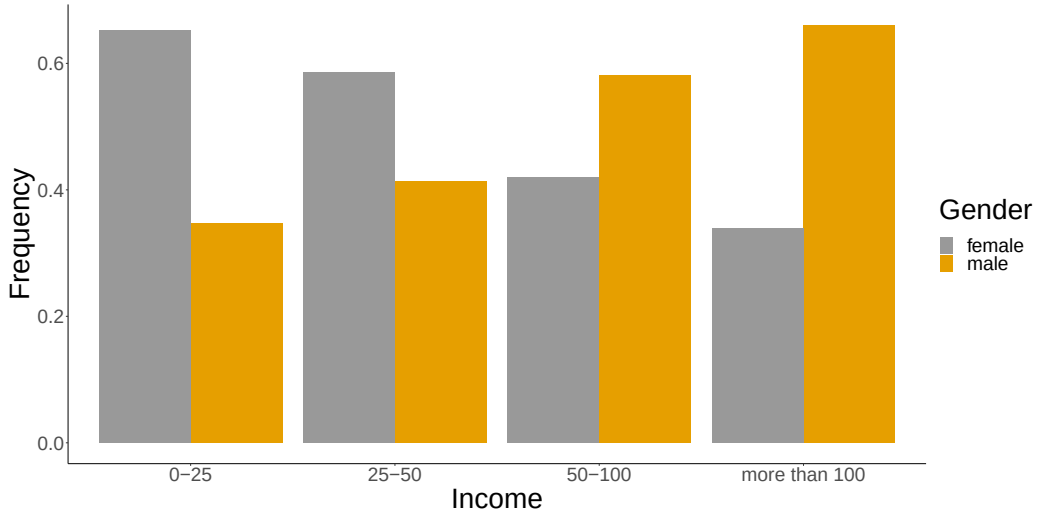
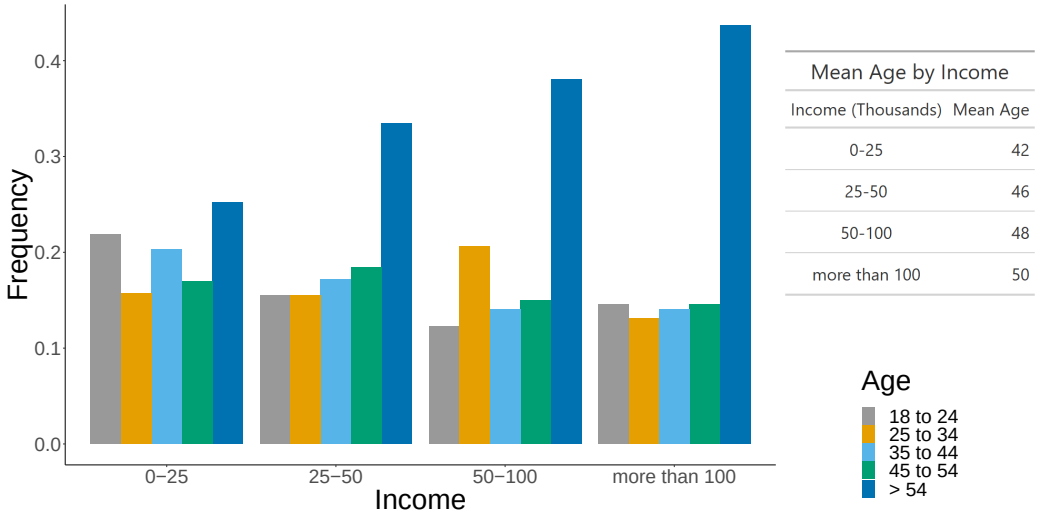


Figure 5: each individual's average response to all 0-10 pain questions after removing protest answers



Mean Age by Income

Income (Thousands)	Mean Age
0-25	42
25-50	46
50-100	48
more than 100	50

Figure 6

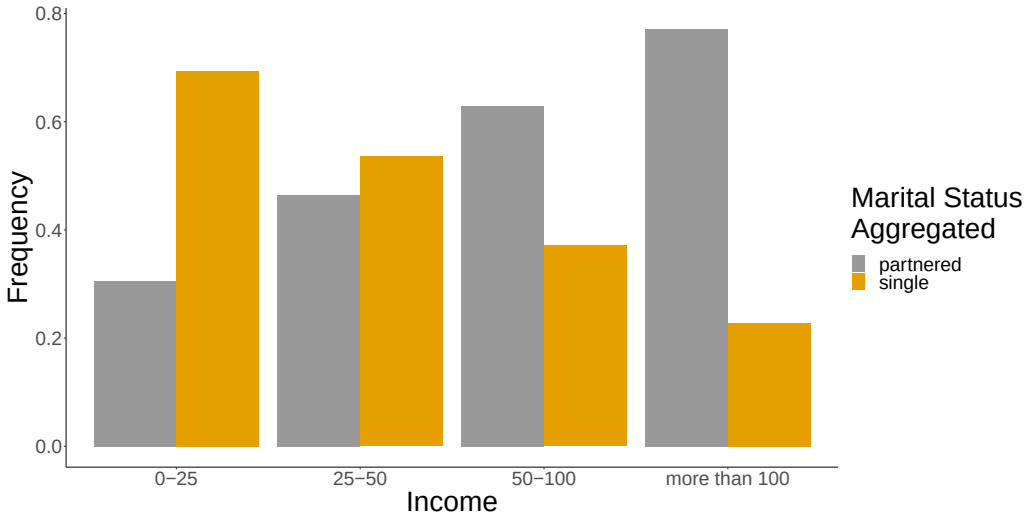


Figure 7: Partnered includes "Living with a partner in a long term relationship" and married, while single is "Single, never married", Widowed, Divorced, Separated.

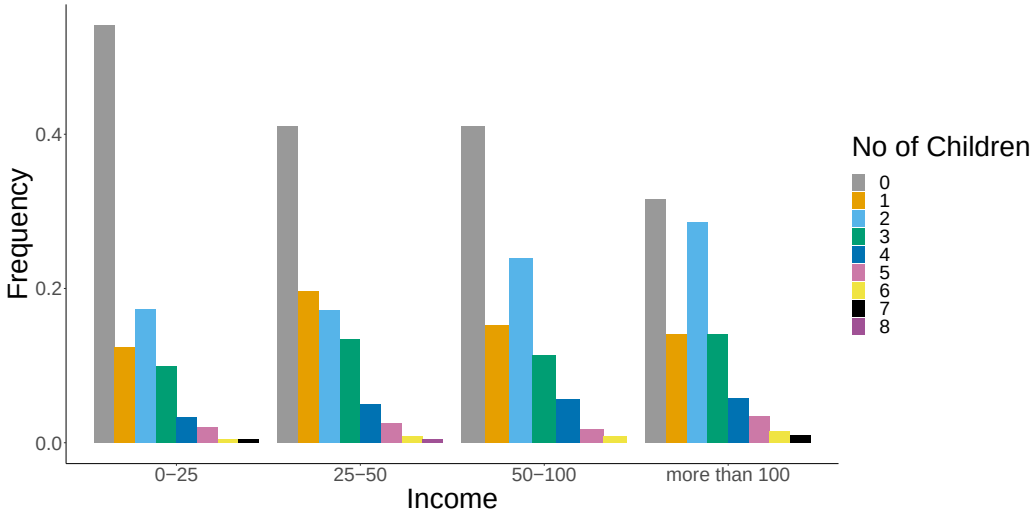


Figure 8: 8 was the max reported in my survey

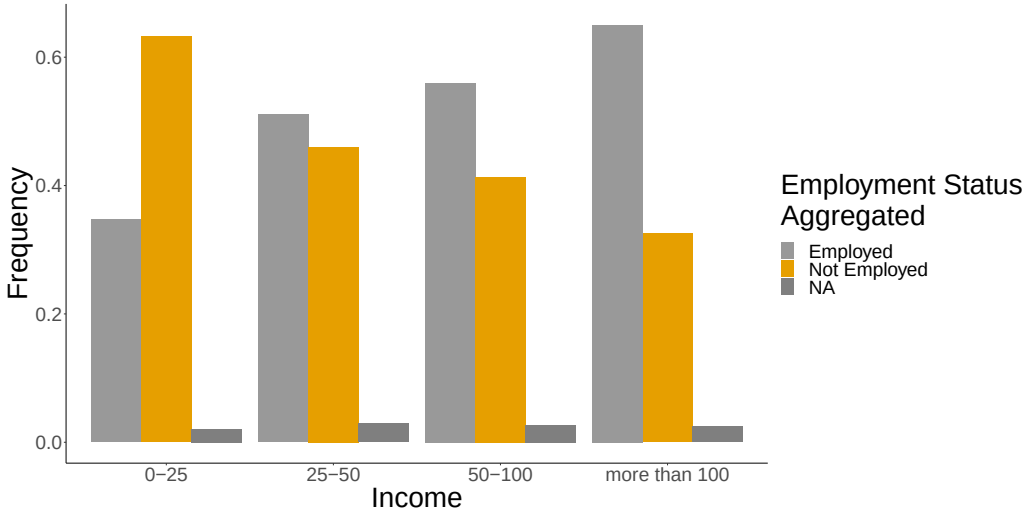


Figure 9

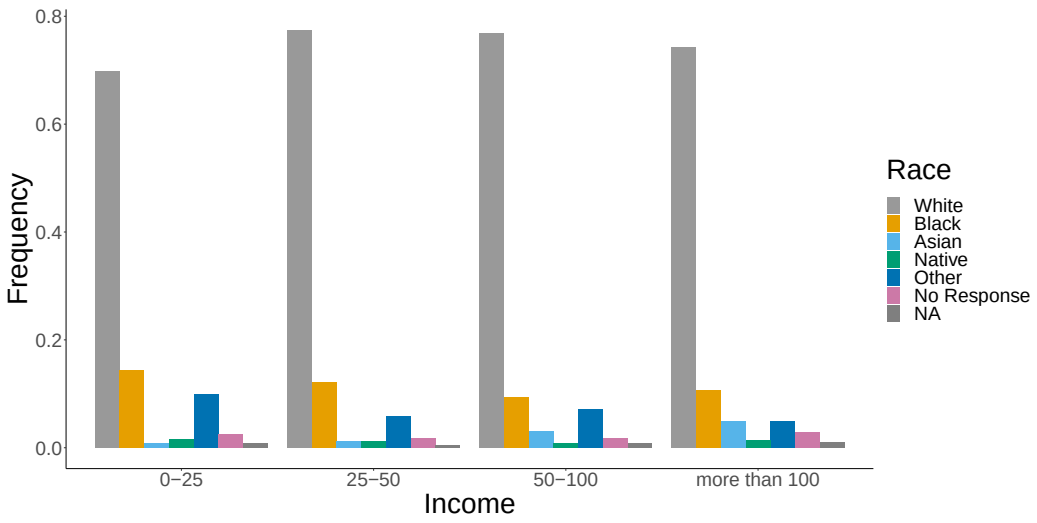


Figure 10

Figure 11

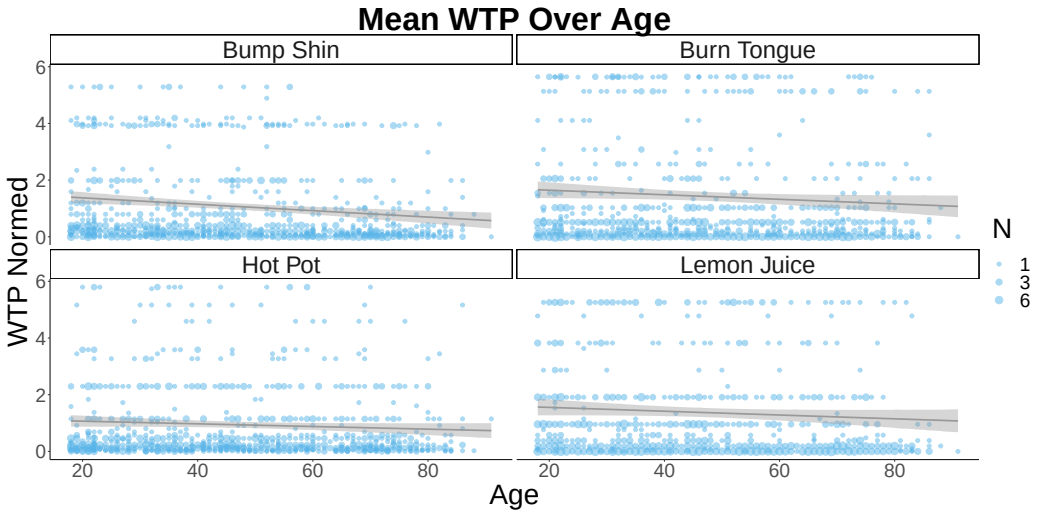
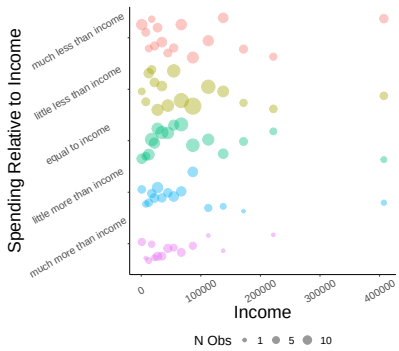
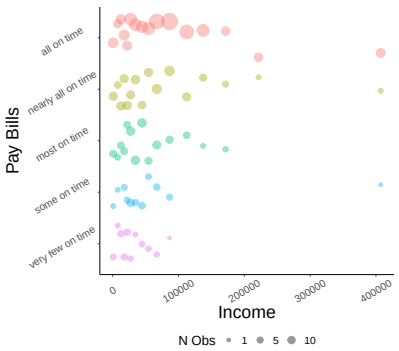
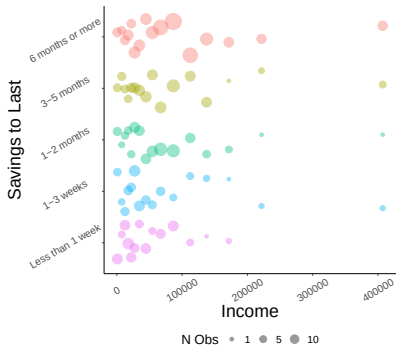
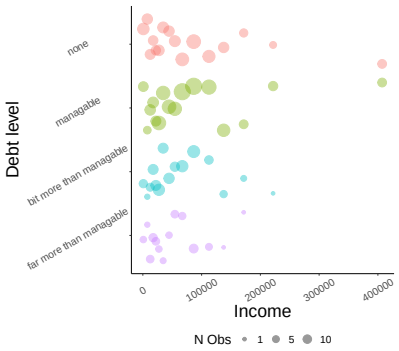

Mean WTP Over Age
Bump Shin
Burn Tongue
Hot Pot
Lemon Juice
WTP Normed
Age
N
1
3
6
0
2
4
6
20
40
60
80

Figure 12


Spending Relative to Income
much less than income
little less than income
equal to income
little more than income
much more than income
0
100000
200000
300000
400000
Income
N Obs
1
5
10


Pay Bills
all on time
nearly all on time
most on time
some on time
very few on time
0
100000
200000
300000
400000
Income
N Obs
1
5
10


Savings to Last
6 months or more
3–5 months
1–2 months
1–3 weeks
Less than 1 week
0
100000
200000
300000
400000
Income
N Obs
1
5
10


Debt level
none
managable
bit more than managable
far more than managable
0
100000
200000
300000
400000
Income
N Obs
1
5
10

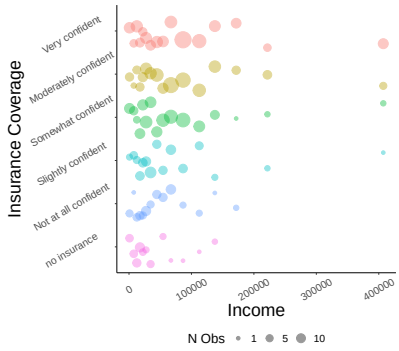

Insurance Coverage
Very confident
Moderately confident
Somewhat confident
Slightly confident
Not at all confident
no insurance
0
100000
200000
300000
400000
Income
N Obs
1
5
10

Figure 13: The answers correspond to questions in appendix

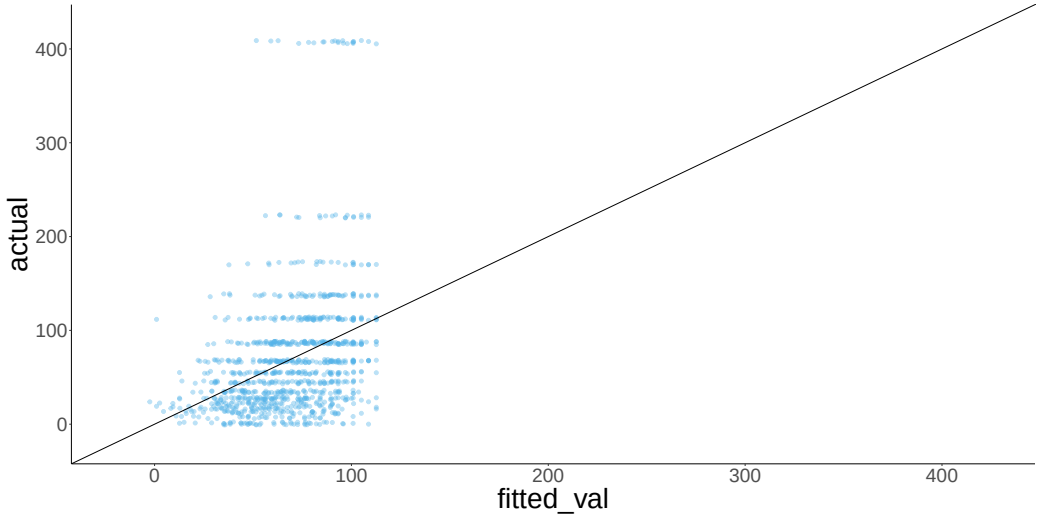
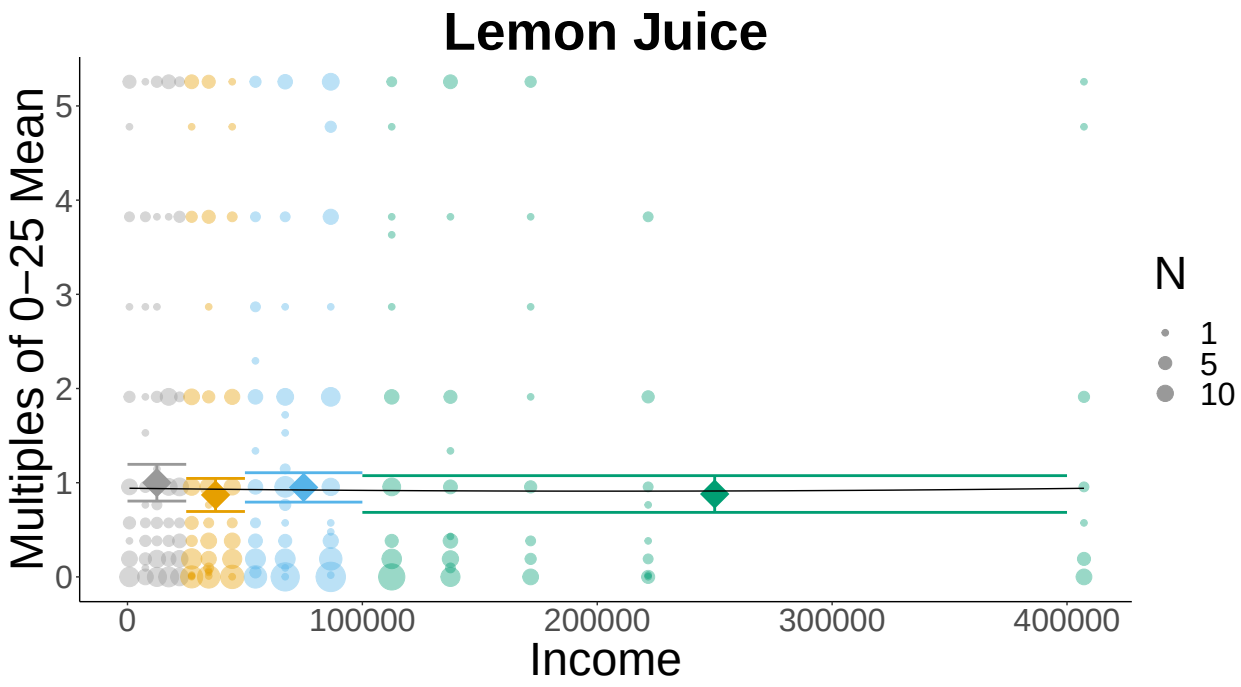


Figure 14

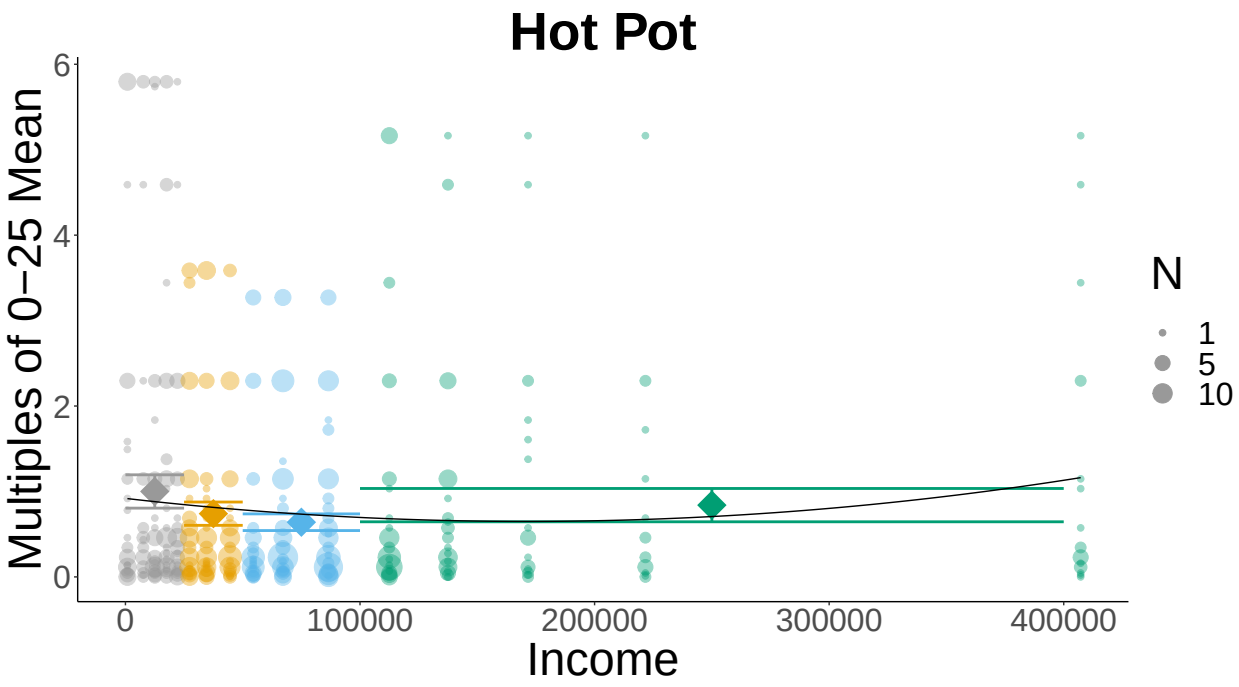


*0−25 mean before normalization was 5.23
*Black fitted line is a cubic polynomial

Quadratic Regression		
Term	Estiamte	Std Error
Intercept	0.94	0.088
Income (1000's)	−0.00030	0.0016
Income^2 (1000's)	0.00000074	0.0000045

Group Means		
Income	Mean	Std Error
0-25	1.00	0.10
25-50	0.87	0.09
50-100	0.95	0.08
more than 100	0.88	0.10

Figure 15

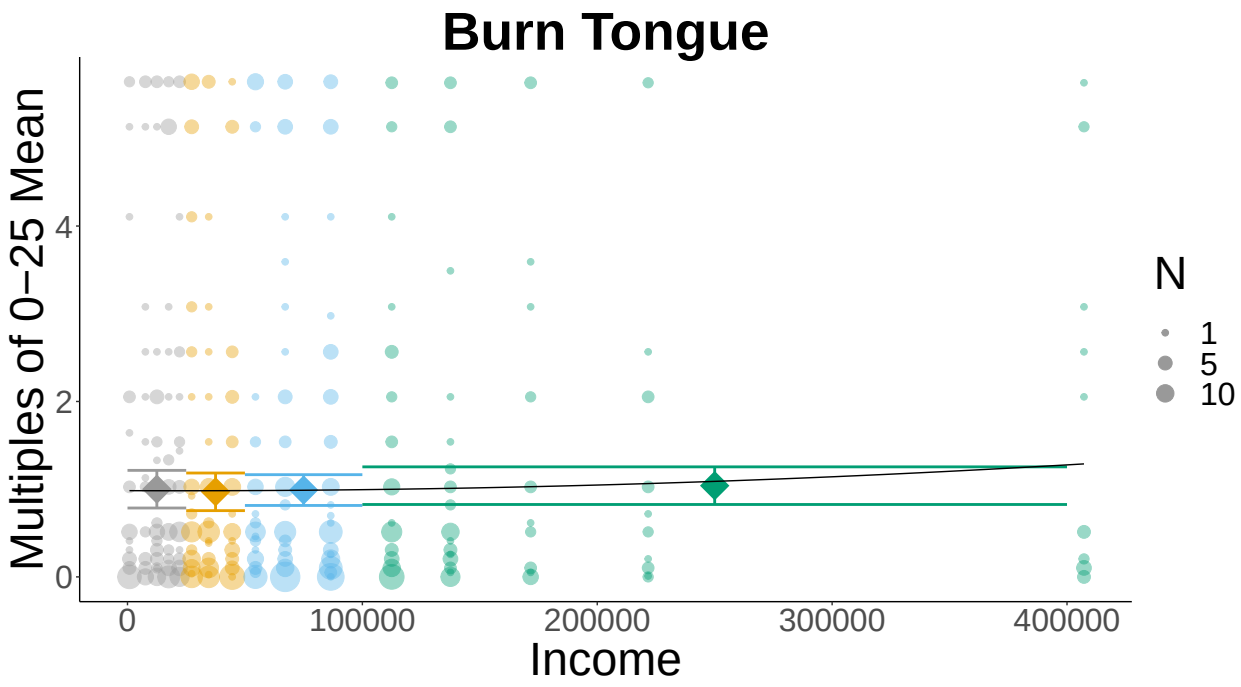


**0−25 mean before normalization was 43.57*
**Black fitted line is a cubic polynomial*

Quadratic Regression		
Term	Estiamte	Std Error
Intercept	0.92	0.073
Income (1000's)	−0.0031	0.0013
Income^2 (1000's)	0.0000092	0.0000037

Group Means		
Income	Mean	Std Error
0-25	1.00	0.10
25-50	0.74	0.07
50-100	0.64	0.05
more than 100	0.84	0.10

Figure 16



*0−25 mean before normalization was 9.74
*Black fitted line is a cubic polynomial

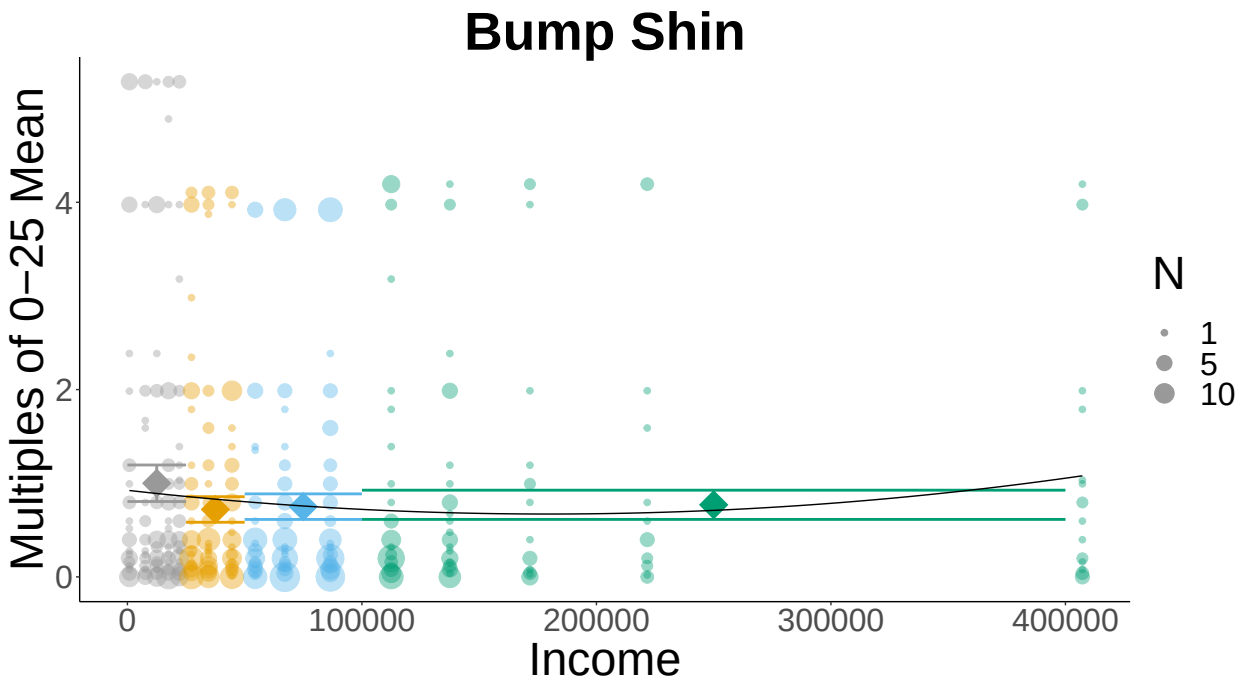
Quadratic Regression

Term	Estiamte	Std Error
Intercept	0.98	0.097
Income (1000's)	−0.000095	0.0018
Income^2 (1000's)	0.0000021	0.0000049

Group Means

Income	Mean	Std Error
0-25	1.00	0.11
25-50	0.97	0.11
50-100	0.99	0.09
more than 100	1.04	0.11

Figure 17

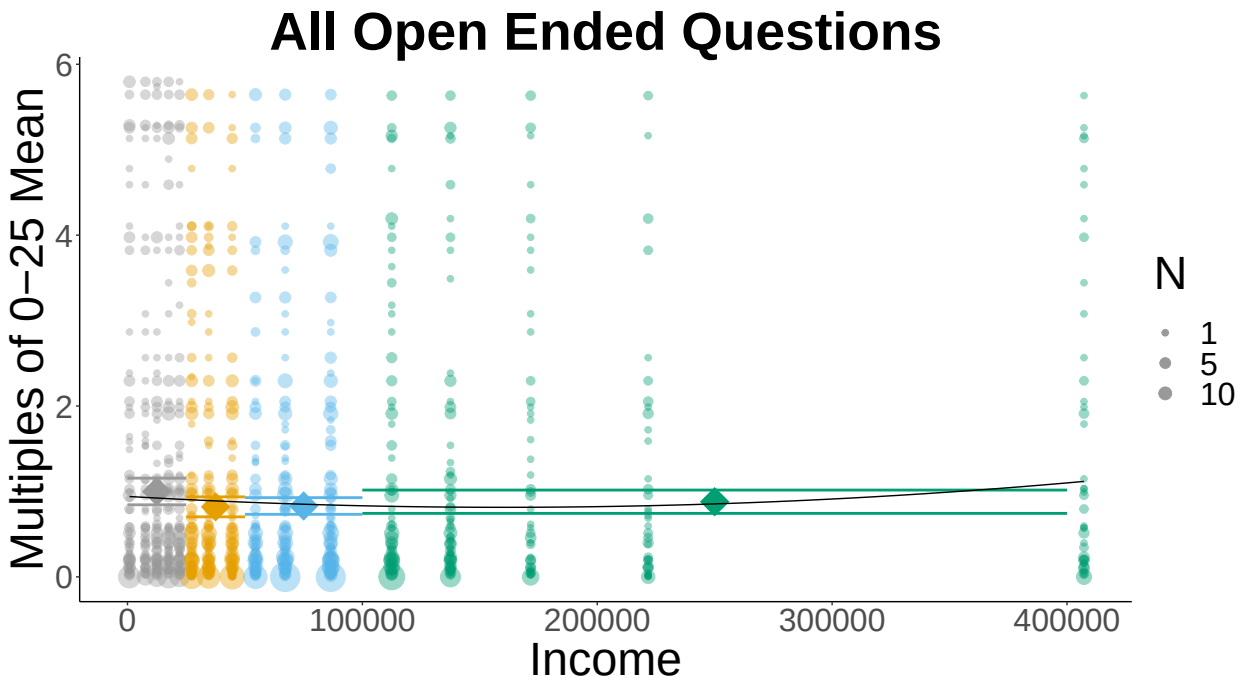


*0−25 mean before normalization was 25.16
*Black fitted line is a cubic polynomial

Quadratic Regression		
Term	Estiamte	Std Error
Intercept	0.92	0.077
Income (1000's)	−0.0028	0.0014
Income^2 (1000's)	0.0000079	0.0000039

Group Means		
Income	Mean	Std Error
0-25	1.00	0.10
25-50	0.72	0.07
50-100	0.75	0.07
more than 100	0.77	0.08

Figure 18



*0−25 mean before normalization was 43.57
*Black fitted line is a cubic polynomial

Quadratic Regression		
Term	Estiamte	Std Error
Intercept	0.94	0.042
Income (1000's)	−0.0016	0.00078
Income^2 (1000's)	0.0000050	0.0000021

Group Means		
Income	Mean	Std Error
0-25	1.00	0.08
25-50	0.82	0.06
50-100	0.83	0.05
more than 100	0.88	0.07

Figure 19

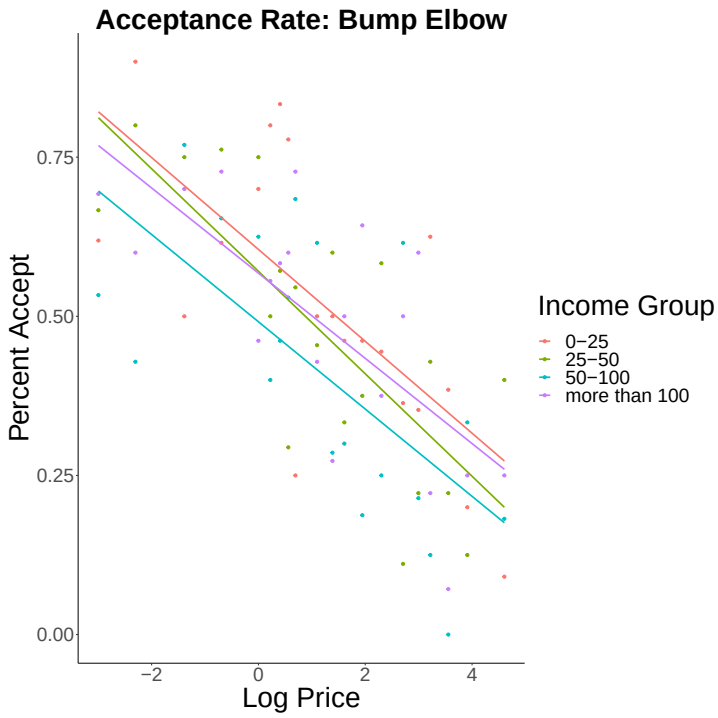


Figure 20

Figure 21

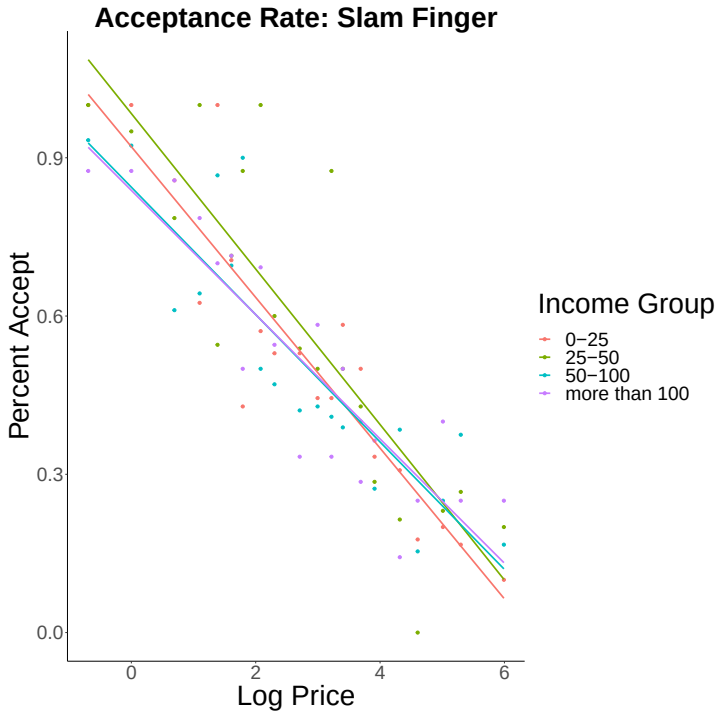


Figure 22

Appendix D Maximum Likelihood Identification

equations 7 and 1 lead to the following theorem specifying the identification of the parameters. Let $\theta = (\boldsymbol{\beta}, \boldsymbol{\phi'})$ be the full set of paremeters in the model.

Theorem 5. *If the conditions in assumption 3 and 4 and definition 1 hold and we also have that the matrix* $[\mathbb{X} \quad \mathbb{M}]$ *is full rank, then the ratio of any two parameters in* θ *is identified. If we normalize the marginal utility of income for the lowest income group to one, that is* $\phi'_1 = 1$*, than the remaining parameters in* θ *are identified.*

To prove this Suppose there exists a $\theta^* \neq \theta$ s.t. $\mathbb{E}[\mathbb{P}^r|\theta^*] = \mathbb{E}[\mathbb{P}^r|\theta]$. This implies

$$\mathbb{X}_i\boldsymbol{\beta}^* \oslash \mathbb{M}\phi'^* = \mathbb{X}_i\boldsymbol{\beta} \oslash \mathbb{M}\phi' \tag{32}$$

or that

$$\frac{\beta_1^* + X_i\boldsymbol{\beta}^*}{\sum_{k=1}^b \mathbb{1}_{ik}\phi_k'^*} = \frac{\beta_1 + X_i\boldsymbol{\beta}}{\sum_{k=1}^b \mathbb{1}_{ik}\phi_k'} \quad \forall \quad i \tag{33}$$

Now it is true that $\theta = \alpha\theta^*$, where α is any constant, satisfies the condition since α cancels out in the numerator and denominator. Once we have normalized $\phi_1 = 1$, however, α no longer appears in the denominator for i's in income group 1 and so does not cancel.

If $\phi_2' = .5$, for example, it implies the marginal utility of a dollar for income group 2 is half that of group 1.

With the $\phi_1 = 1$ normalization, any change to a parameter in the numerator would alter the expected reservation price for those in group one and violate the equality in equation 33, assuming, as in a regression, that $\mathbb{X}$ is full rank. Any change to the other marginal utility parameters ϕ_k' could be cancelled out for that group by appropriately scaling the numerator, but, since the marginal utility of group one is fixed and they share the numerator parameters, this would again change the expected reservation price for group 1 and violate 33. Thus, there does not exist a $\theta^* \neq \theta$ satisfying the condition.

Can this same identification strategy be used for any good? No, the marginal utility of income cannot be identified from just any reservation price. Recall the assumption that $\epsilon_i \perp\!\!\!\perp m_i$ and notice that income does not enter the utility function for pain relief and the factors impacting pain relief to not impact the marginal utility of income. If either of these appeared in both the numerator and denominator, we would not not be able to uniquely identify the parameters.

Appendix E Full Centiment Survey

Centiment Survey

The instructions where first written alone on an initial page and required respondents to wait five seconds before continuing.

Instructions copy ***Instructions Repeated (if needed for reference):***

Imagine yourself in each of the following situations.

Consider if each situation would be painful for you and if yes, how painful it would be.

Then, enter the MOST you would pay in U.S. dollars to completely and immediately eliminate any pain caused by the situation, as if the event described never happened.

Q1 Imagine you have a minor cut on your finger and you accidentally get lemon juice in the wound.

- ○ $ _____

Q2 Imagine you pick up a hot pot by accidentally grabbing its equally hot handles.

- ○ $ _____

Q3 Imagine you shake hands with someone who has a normal grip.

- ○ $ _____

Q4 Imagine you burn your tongue on a very hot drink.

○ $ _____

Q5 Imagine you bump your shin badly on a hard edge, for example, on the edge of a glass coffee table.

○ $ _____

End of Block: WTP Pain

Start of Block: bump your elbow

Q6 Imagine you bump your elbow on the edge of a table (“funny bone”). Would you pay $X to completely and immediately eliminate any pain caused by this situation, as if it never happened?

○ Yes, I would pay $X

○ No, I would not pay $X

NOTE: X is chosen Randomly from (0.05 0.10 0.25 0.50 1 1.25 1.50 1.75 2 3 4 5 7 10 15 20 25 35 50 100) and a single value is displayed

End of Block: bump your elbow

Start of Block: Bite cheek

Q7 Imagine you accidentally bite your tongue or cheek badly while eating. Would you pay $X to completely and immediately eliminate any pain caused by this situation, as if it never happened?

- ○ Yes, I would pay $X
- ○ No, I would not pay $X

NOTE: X is chosen Randomly from (0.10 0.25 0.50 0.75 1 2 3 4 5 6 8 10 15 20 25 35 50 75 100 200) and a single value is displayed

End of Block: Bite cheek

Start of Block: Trap finger drawer

Q8 Imagine you slam your finger in a drawer. Would you pay $X to completely and immediately eliminate any pain caused by this situation, as if it never happened?

- ○ Yes, I would pay $X
- ○ No, I would not pay $X

NOTE: X is chosen Randomly from 0.5 1 2 3 4 5 6 8 10 15 20 25 30 40 50 75 100 150 200 400) and a single value is displayed

End of Block: Trap finger drawer

Start of Block: PSQ Match

Q9 Imagine yourself in each of the following situations. Decide if each situation would be painful for you and if yes, how painful it would be. Let 0 stand for no pain; 1 is just noticeable pain and 10 is the most severe pain that you can imagine.

	0 (no pain) (1)	1 (2)	2 (3)	3 (4)	4 (5)	5 (6)	6 (7)	7 (8)	8 (9)	9 (10)	10 (most severe pain imaginable) (11)
Imagine you slam your finger in a drawer. (1)	○	○	○	○	○	○	○	○	○	○	○
Imagine you accidentally bite your tongue or cheek badly while eating. (2)	○	○	○	○	○	○	○	○	○	○	○
Imagine you shake hands with someone who has a normal grip. (4)	○	○	○	○	○	○	○	○	○	○	○
Imagine you pick up a hot pot by accidentally grabbing its equally hot handles. (3)	○	○	○	○	○	○	○	○	○	○	○
Imagine you burn your tongue on a very hot drink. (5)	○	○	○	○	○	○	○	○	○	○	○
To ensure your full attention, please select 9 (8)	○	○	○	○	○	○	○	○	○	○	○

Imagine you bump your elbow on the edge of a table (“funny bone”). (6)	○	○	○	○	○	○	○	○	○	○	○
Imagine you bump your shin badly on a hard edge, for example, on the edge of a glass coffee table. (7)	○	○	○	○	○	○	○	○	○	○	○

Skip To: End of Block If Imagine yourself in each of the following situations. Decide if each situation would be painful f... != To ensure your full attention, please select 9 [9]

End of Block: PSQ Match

Start of Block: Demographics

Q10 Which of the following best describes your status?

- ○ Married
- ○ Divorced
- ○ Widowed
- ○ Separated
- ○ Single, never married
- ○ Living with a partner in a long term relationship

*

Q11 How many children do you have?

_____

Q13 What is the highest level of school you have completed or the highest degree you have received?

- Less than high school
- High School Graduate- high school DIPLOMA or the equivalent (For example: GED)
- Some college but no degree
- Occupational/vocational program
- Associate degree in college
- Bachelor's degree (For example: BA, AB, BS)
- Master's degree (For example: MA, MS, MEng, MEd, MSW, MBA)
- Professional School Degree (For example: MD,DDS,DVM,LLB,JD)
- Doctorate degree (For example: PhD, EdD)

Q14 What is your current employment status? Mark all that apply.

- [ ] Employed for wages full time
- [ ] Employed for wages part time
- [ ] Self-employed
- [ ] Out of work and looking for work
- [ ] Out of work but not currently looking for work
- [ ] A home-maker
- [ ] A student
- [ ] Military
- [ ] Retired
- [ ] Unable to work

End of Block: Demographics

Start of Block: Financial Health FHN

Title For the following questions, think of a HOUSEHOLD as all people related by birth, marriage, or adoption and residing together. If you live alone, or do not consider anyone else to be a member of your HOUSEHOLD, please answer these questions as an individual.

Q15 Which category represents the total combined income of all members of your HOUSEHOLD during the past 12 months? This includes money from jobs, net income from business, farm or rent, pensions, dividends, interest, social security payments and any other money income received by members of your HOUSEHOLD who are 15 years of age or older.

- ○ Less than $5,000
- ○ $5,000 to $9,999
- ○ $10,000 to $14,999
- ○ $15,000 to $19,999
- ○ $20,000 to $24,999
- ○ $25,000 to $29,999
- ○ $30,000 to $39,999
- ○ $40,000 to $49,999
- ○ $50,000 to $59,999
- ○ $60,000 to $74,999
- ○ $75,000 to $99,999
- ○ $100,000 to $124,999
- ○ $125,000- $149,999
- ○ $150,000 - $199,999
- ○ $200,000 to $249,999
- ○ $250,000 and over

Q16 Which of the following statements best describes how your household's total spending compared to total income over the last 12 months?

- ○ Spending was much less than income
- ○ Spending was a little less than income
- ○ Spending was about equal to income
- ○ Spending was a little more than income
- ○ Spending was much more than income

Q17 Which of the following statements best describes how your household has paid its bills over the last 12 months? My household has been financially able to:

- ○ Pay all our bills on time
- ○ Pay nearly all our bills on time
- ○ Pay most of our bills on time
- ○ Pay some of our bills on time
- ○ Pay very few of our bills on time

Q18 At your current level of spending, how long could you and your household afford to cover expenses, if you had to live on only the money you have readily available, without withdrawing money from retirement accounts or borrowing?

- ○ 6 months or more
- ○ 3-5 months
- ○ 1-2 months
- ○ 1-3 weeks
- ○ Less than 1 week

Q19 Thinking about all of your household's current debts, including mortgages, bank loans, student loans, money owed to people, medical debt, past-due bills, and credit card balances that are carried over from prior months...

As of today, which of the following statements describes how manageable your household debt is?

- ○ Do not have any debt
- ○ Have a manageable amount of debt
- ○ Have a bit more debt than is manageable
- ○ Have far more debt than is manageable

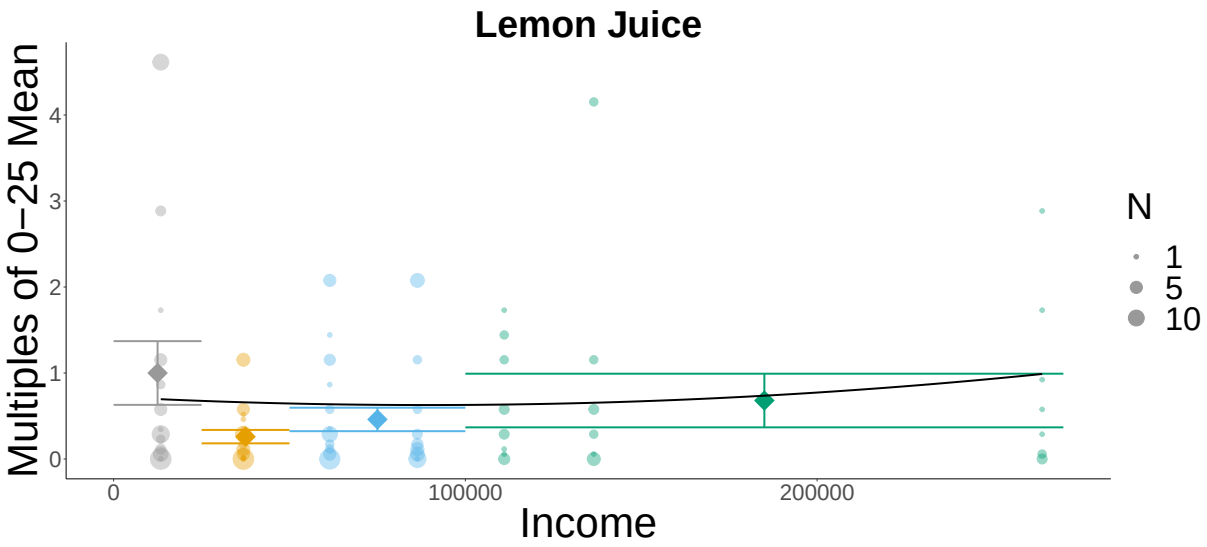
Q20 Thinking about all of the types of insurance you and others in your household currently might have, including health insurance, vehicle insurance, home or rental insurance, life insurance, and disability insurance...

How confident are you that those insurance policies will provide enough support in case of an emergency?

- ○ Very confident
- ○ Moderately confident
- ○ Somewhat confident
- ○ Slightly confident
- ○ Not at all confident
- ○ No one in my household has any insurance

End of Block: Financial Health FHN

Appendix F Pollfish results

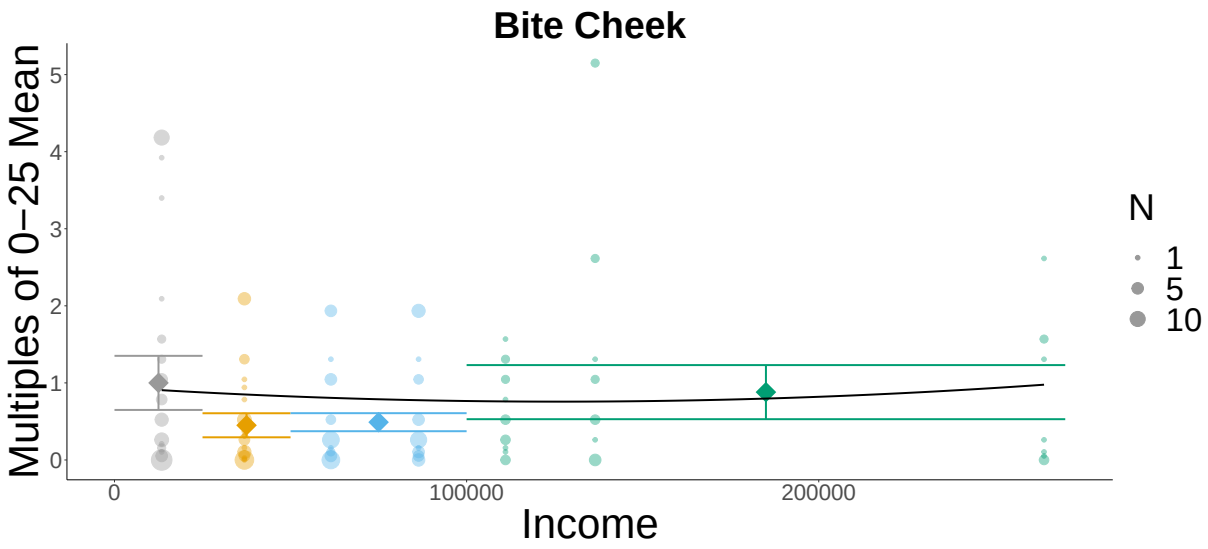


*0−25 mean before normalization was 17.33
*Black fitted line is a cubic polynomial

Lemon Juice					
Income	Mean	se	Std Dev	Std Dev Norm	Median Norm
0-25	1.00	0.18	1.59	1.00	1.0
25-50	0.26	0.05	0.35	0.22	0.4
50-100	0.46	0.07	0.70	0.44	0.4
more than 100	0.68	0.16	1.00	0.63	1.0

MLE Lemon Juice			
label	est	se	P_ratio
male	1.42	3.39	NA
psq	4.91	1.20	NA
intercept	−0.70	4.74	NA
sigma	26.11	2.00	NA
0-25	**1.00**	NA	**1.0000000**
25-50	**4.51**	0.49	**0.2219545**
50-100	**2.21**	0.22	**0.4524891**
more than 100	**1.52**	0.19	**0.6570436**
mean psq control			

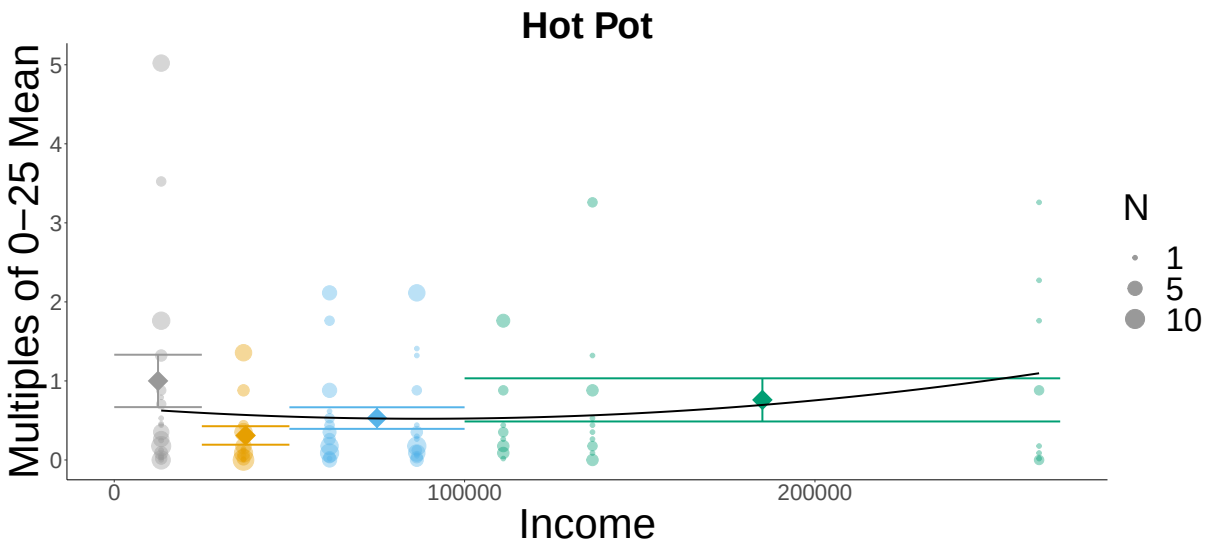
Figure 23: imagine you have a minor cut on your finger and you inadvertently get lemon juice in the wound



Bite Cheek					
Income	Mean	se	Std Dev	Std Dev Norm	Median Norm
0-25	1.00	0.17	1.47	1.00	1.0
25-50	0.45	0.08	0.65	0.45	0.4
50-100	0.49	0.07	0.64	0.44	1.0
more than 100	0.88	0.18	1.21	0.82	2.0

MLE Bite Cheek			
label	est	se	P_ratio
male	5.05	3.42	NA
psq	5.34	1.20	NA
intercept	−1.80	4.75	NA
sigma	26.22	1.98	NA
0-25	**1.00**	NA	**1.0000000**
25-50	**2.29**	0.24	**0.4374243**
50-100	**2.19**	0.21	**0.4575944**
more than 100	**1.16**	0.14	**0.8607178**
mean psq control			

Figure 24: imagine you accidentally bite your tongue or cheek badly while eating

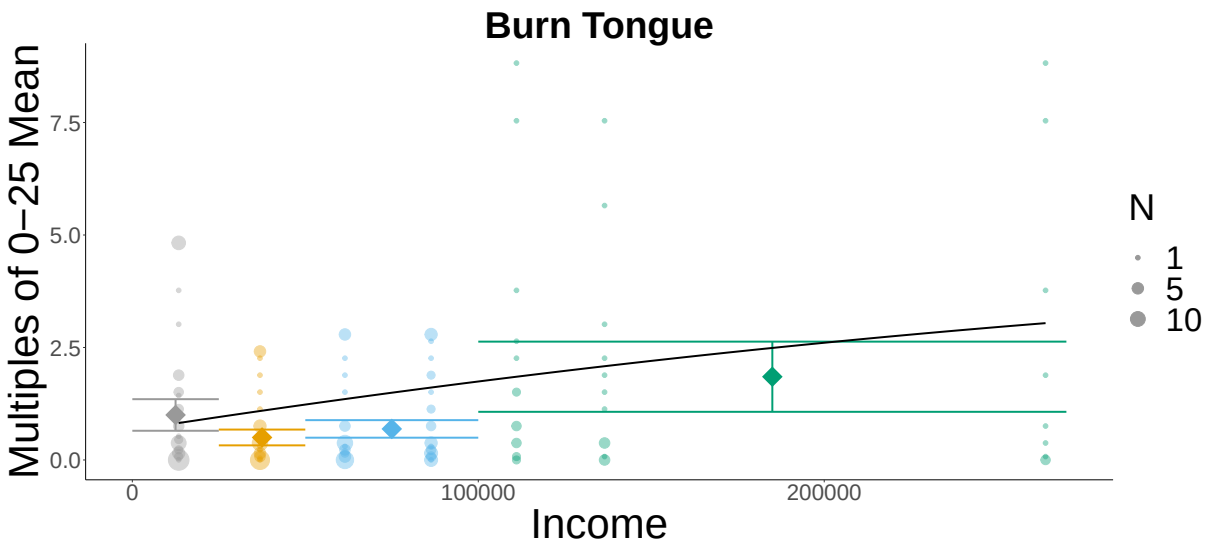


*0−25 mean before normalization was 56.76
*Black fitted line is a cubic polynomial

Hot Pot					
Income	Mean	se	Std Dev	Std Dev Norm	Median Norm
0-25	1.00	0.17	1.52	1.00	1.00
25-50	0.31	0.06	0.44	0.29	0.33
50-100	0.53	0.07	0.71	0.46	0.67
more than 100	0.76	0.15	0.93	0.61	1.33

MLE Hot Pot			
label	est	se	P_ratio
male	12.86	10.49	NA
psq	14.20	3.65	NA
intercept	3.24	14.62	NA
sigma	80.50	6.10	NA
0-25	**1.00**	NA	**1.0000000**
25-50	**3.45**	0.37	**0.2899111**
50-100	**2.05**	0.20	**0.4885869**
more than 100	**1.51**	0.18	**0.6608440**
mean psq control			

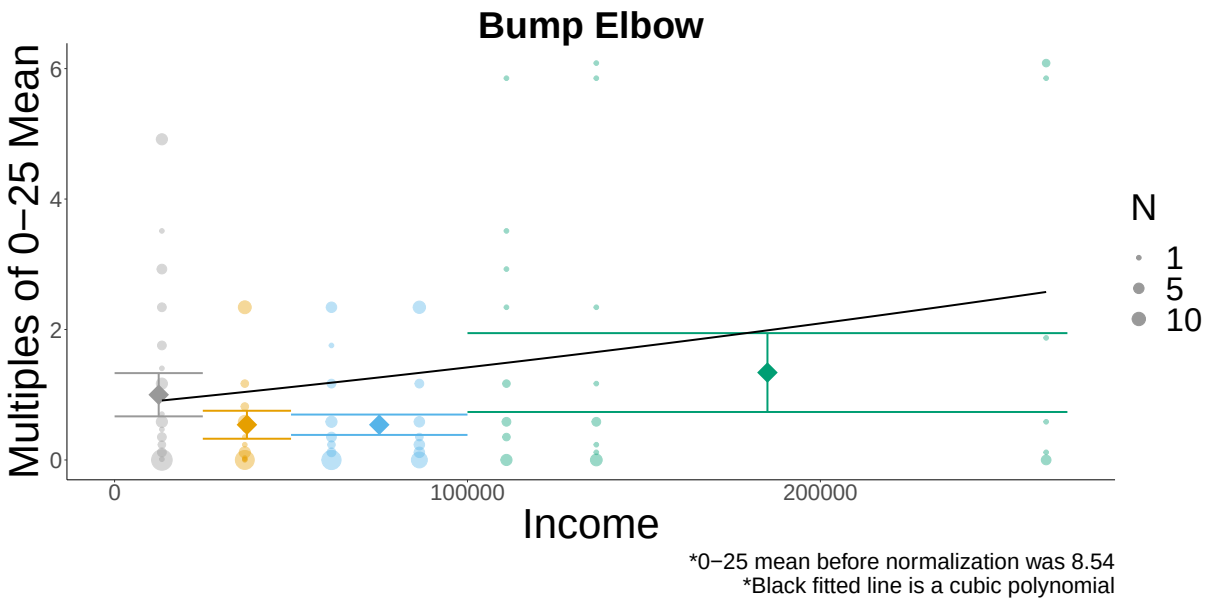
Figure 25: imagine you pick up a hot pot by inadvertently grabbing its equally hot handles



Burn Tongue					
Income	Mean	se	Std Dev	Std Dev Norm	Median Norm
0-25	1.00	0.18	1.54	1.00	1.0
25-50	0.50	0.10	0.74	0.48	0.4
50-100	0.69	0.10	0.95	0.61	0.7
more than 100	1.85	0.40	2.60	1.69	2.0

MLE Burn Tongue			
label	est	se	P_ratio
male	6.96	2.50	NA
psq	3.75	0.88	NA
intercept	−2.46	3.45	NA
sigma	19.01	1.44	NA
0-25	**1.00**	NA	**1.0000000**
25-50	**2.12**	0.23	**0.4716562**
50-100	**1.56**	0.15	**0.6407617**
more than 100	**0.58**	0.07	**1.7384238**
mean psq control			

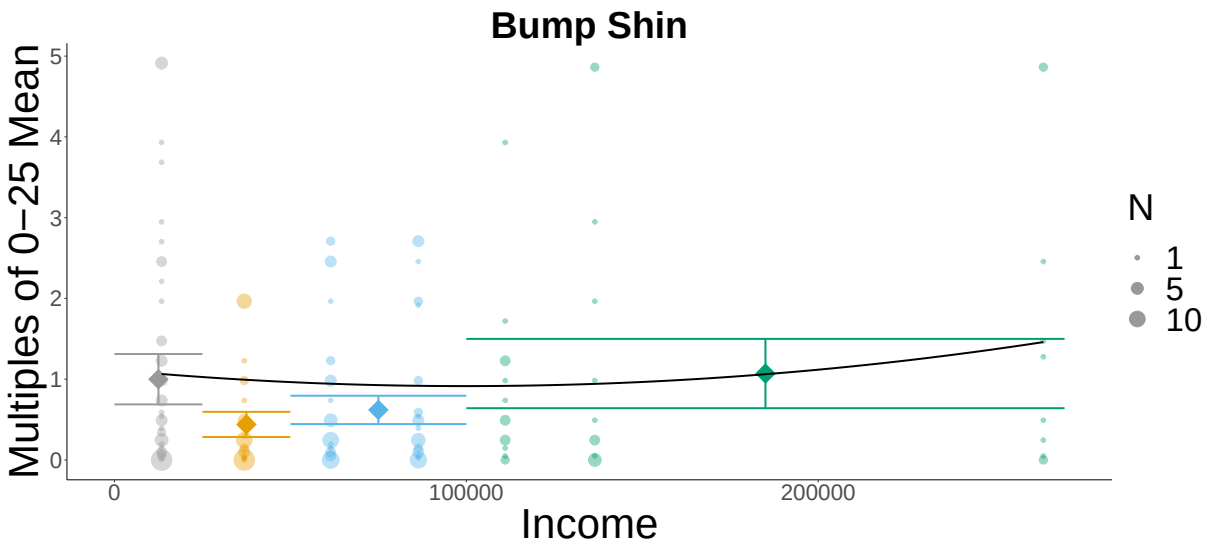
Figure 26: imagine you burn your tongue on a very hot drink.



Bump Elbow					
Income	Mean	se	Std Dev	Std Dev Norm	Median Norm
0-25	1.00	0.18	1.50	1.00	1.0
25-50	0.54	0.11	0.83	0.55	0.5
50-100	0.54	0.09	0.81	0.54	0.5
more than 100	1.34	0.30	2.06	1.38	1.5

MLE Bump Elbow			
label	est	se	P_ratio
male	1.71	1.58	NA
psq	2.50	0.56	NA
intercept	−1.52	2.19	NA
sigma	12.11	0.92	NA
0-25	**1.00**	NA	**1.0000000**
25-50	**1.86**	0.20	**0.5375826**
50-100	**1.81**	0.18	**0.5511262**
more than 100	**0.71**	0.09	**1.4090528**
mean psq control			

Figure 27: imagine you bump your elbow on the edge of a table ("funny bone").

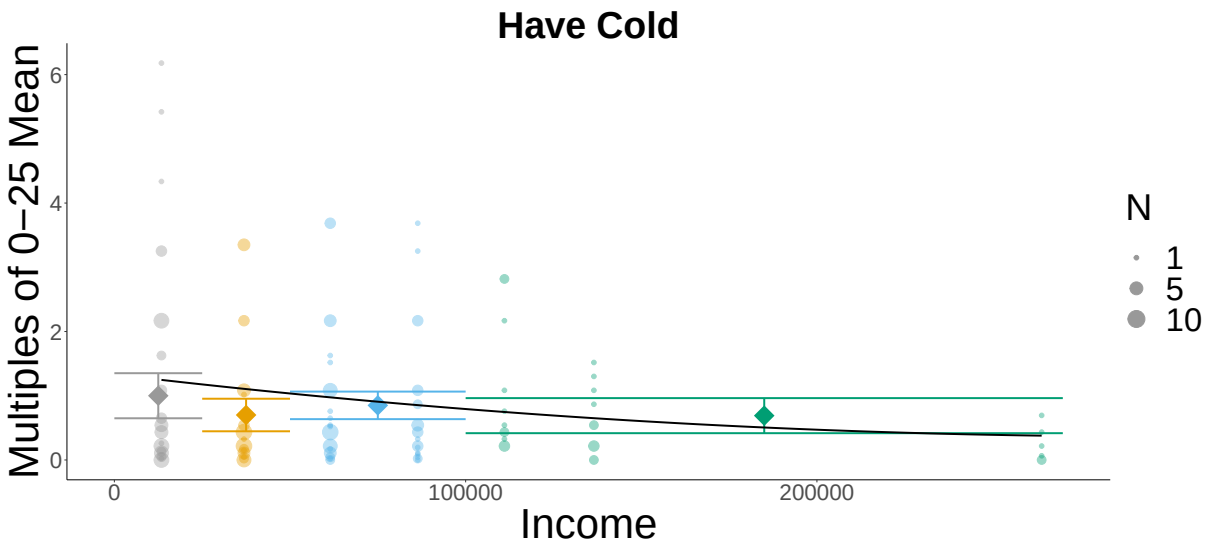


*0−25 mean before normalization was 20.35
*Black fitted line is a cubic polynomial

Bump Shin					
Income	Mean	se	Std Dev	Std Dev Norm	Median Norm
0-25	1.00	0.17	1.41	1.00	1.00
25-50	0.44	0.08	0.66	0.47	0.40
50-100	0.62	0.09	0.85	0.60	0.67
more than 100	1.07	0.23	1.50	1.06	1.33

MLE Bump Shin			
label	est	se	P_ratio
male	10.52	3.61	NA
psq	4.48	1.23	NA
intercept	−0.20	4.92	NA
sigma	27.17	2.05	NA
0-25	**1.00**	NA	**1.0000000**
25-50	**2.16**	0.23	**0.4628516**
50-100	**1.63**	0.16	**0.6149244**
more than 100	**0.92**	0.11	**1.0864257**
mean psq control			

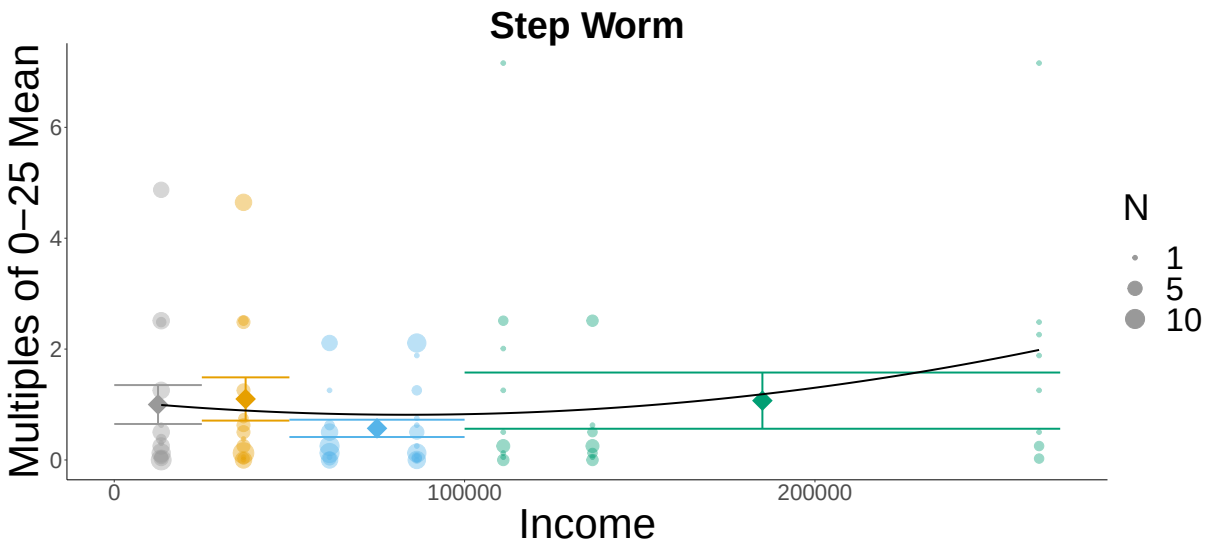
Figure 28: imagine you bump your shin badly on a hard edge, for example, on the edge of a glass coffee table.



Have Cold					
Income	Mean	se	Std Dev	Std Dev Norm	Median Norm
0-25	1.00	0.18	1.38	1.00	1.00
25-50	0.70	0.13	0.95	0.69	0.87
50-100	0.85	0.12	0.99	0.72	1.00
more than 100	0.69	0.14	0.78	0.56	1.00

MLE Have Cold			
label	est	se	P_ratio
male	−0.43	9.19	NA
psq	4.78	2.95	NA
intercept	31.65	12.26	NA
sigma	61.41	5.27	NA
0-25	**1.00**	NA	**1.0000000**
25-50	**1.45**	0.17	**0.6903644**
50-100	**1.32**	0.15	**0.7562644**
more than 100	**1.68**	0.24	**0.5951526**
mean psq control			

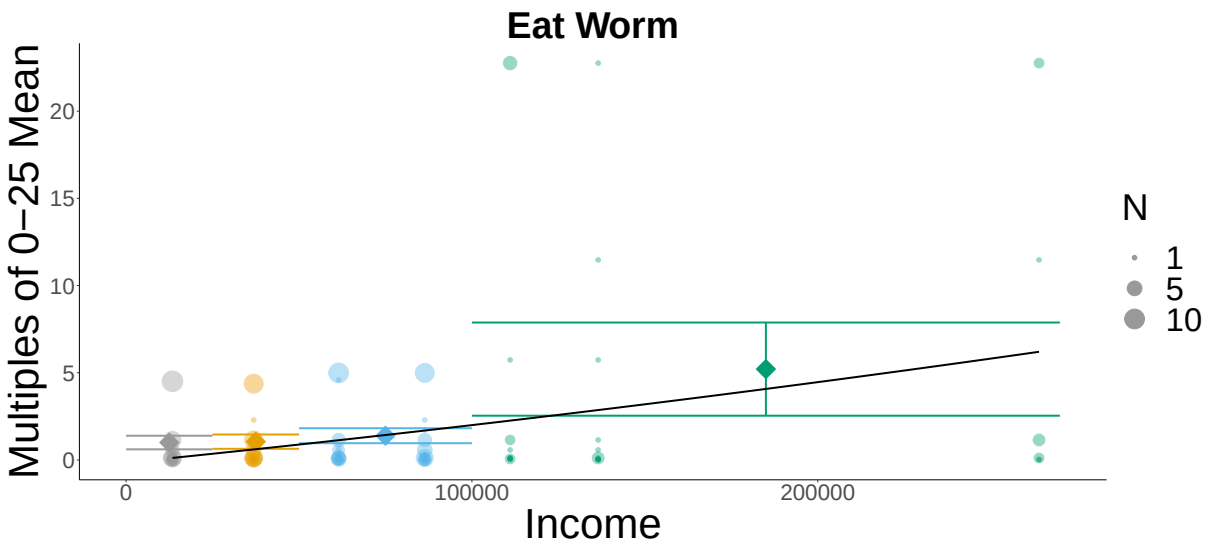
Figure 29: imagine you have a sore throat and cough which will last a week with no other complications. drugs will not relieve the symptoms.



Step Worm					
Income	Mean	se	Std Dev	Std Dev Norm	Median Norm
0-25	1.00	0.17	1.46	1.00	1.00
25-50	1.10	0.20	1.55	1.06	1.25
50-100	0.57	0.08	0.76	0.52	1.00
more than 100	1.07	0.26	1.67	1.15	1.00

MLE Step Worm			
label	est	se	P_ratio
male	−10.28	7.45	NA
psq	8.80	2.64	NA
intercept	11.15	10.47	NA
sigma	56.19	4.42	NA
0-25	**1.00**	NA	**1.0000000**
25-50	**0.94**	0.11	**1.0640200**
50-100	**1.90**	0.19	**0.5271882**
more than 100	**0.87**	0.11	**1.1440865**
mean psq control			

Figure 30: walk barefoot on concrete and step on a dead earthworm.

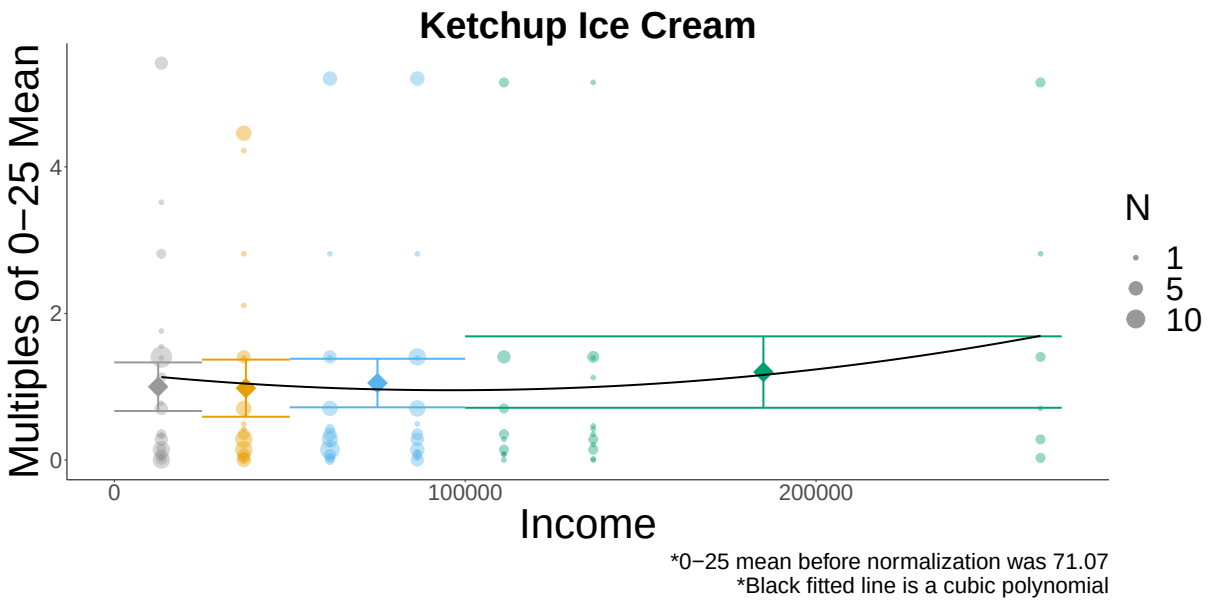


*0−25 mean before normalization was 871.65
*Black fitted line is a cubic polynomial

Eat Worm					
Income	Mean	se	Std Dev	Std Dev Norm	Median Norm
0-25	1.00	0.20	1.65	1.00	1
25-50	1.05	0.21	1.57	0.95	1
50-100	1.39	0.21	2.00	1.21	1
more than 100	5.21	1.34	8.74	5.30	1

MLE Eat Worm			
label	est	se	P_ratio
male	15.35	190.38	NA
psq	221.45	67.95	NA
intercept	54.94	270.47	NA
sigma	1,385.03	116.31	NA
0-25	**1.00**	NA	**1.000000**
25-50	**1.04**	0.12	**0.962051**
50-100	**0.80**	0.08	**1.253803**
more than 100	**0.18**	0.02	**5.438409**
mean psq control			

Figure 31: eat a dead earthworm, 6 inches long.



Ketchup Ice Cream					
Income	Mean	se	Std Dev	Std Dev Norm	Median Norm
0-25	1.00	0.18	1.38	1.00	1.00
25-50	0.98	0.19	1.46	1.05	0.80
50-100	1.05	0.16	1.56	1.13	1.00
more than 100	1.20	0.24	1.60	1.16	1.26

MLE Ketchup Ice Cream			
label	est	se	P_ratio
male	−20.43	13.08	NA
psq	13.85	4.61	NA
intercept	22.47	18.30	NA
sigma	97.00	7.83	NA
0-25	**1.00**	NA	**1.000000**
25-50	**0.97**	0.11	**1.028962**
50-100	**0.90**	0.09	**1.107206**
more than 100	**0.85**	0.11	**1.178982**
mean psq control			

Figure 32: put two tablespoons of ketchup on one cup of vanilla ice cream and eat it.

	Lemon Juice		Bite Cheek		Hot Pot		Burn Tongue		Bump Elbow		Bump Shin	
label	est	se	est	se	est	se	est	se	est	se	est	se
male	1.4248556	3.3938385	5.048771	3.4221608	12.857169	10.4872904	6.9620718	2.50459055	1.7069535	1.57755635	10.5154235	3.6075709
psq	4.9132981	1.2014617	5.337530	1.2025717	14.201043	3.6514827	3.7493992	0.87605874	2.5032895	0.55823489	4.4849073	1.2271831
intercept	-0.7043961	4.7383238	-1.797299	4.7490253	3.235489	14.6205175	-2.4556670	3.44633306	-1.5198520	2.19186640	-0.1995436	4.9161006
sigma	26.1061732	2.0019957	26.215477	1.9792671	80.501951	6.0978881	19.0063680	1.43743405	12.1052190	0.92273439	27.1674595	2.0540339
0-25	**1.0000000**	NA	**1.000000**	NA	**1.000000**	NA	**1.0000000**	NA	**1.0000000**	NA	**1.0000000**	NA
25-50	**4.5054288**	0.4899229	**2.286110**	0.2444016	**3.449333**	0.3709161	**2.1201884**	0.22743035	**1.8601794**	0.20151247	**2.1605195**	0.2329265
50-100	**2.2099982**	0.2182993	**2.185342**	0.2105613	**2.046719**	0.1987606	**1.5606426**	0.15116269	**1.8144664**	0.17828592	**1.6262162**	0.1572062
more than 100	**1.5219689**	0.1867252	**1.161821**	0.1390792	**1.513216**	0.1811029	**0.5752337**	0.06865261	**0.7096966**	0.08613581	**0.9204495**	0.1100030

	Have Cold		Step Worm		Eat Worm		Ketchup Ice Cream	
label	est	se	est	se	est	se	est	se
male	-0.4336552	9.1943315	-10.2760847	7.4511514	15.3547878	190.37874196	-20.4301353	13.07856097
psq	4.7768386	2.9515214	8.8045662	2.6367809	221.4533352	67.94913386	13.8538416	4.60867985
intercept	31.6474190	12.2647548	11.1454741	10.4689092	54.9444072	270.46596055	22.4739752	18.30406629
sigma	61.4140409	5.2724097	56.1932747	4.4198356	1385.0300168	116.30663366	96.9989201	7.83042050
0-25	**1.0000000**	NA	**1.0000000**	NA	**1.0000000**	NA	**1.0000000**	NA
25-50	**1.4485104**	0.1735345	**0.9398320**	0.1059102	**1.0394459**	0.12284045	**0.9718535**	0.11212166
50-100	**1.3222888**	0.1450972	**1.8968559**	0.1894838	**0.7975737**	0.08411142	**0.9031741**	0.09184296
more than 100	**1.6802415**	0.2357622	**0.8740598**	0.1092534	**0.1838773**	0.02418736	**0.8481892**	0.10600972

Average Pain Questions				
Income	Mean	Std Dev	Std Dev Norm	Median Norm
0-25	1.00	1.50	1.00	1.00
25-50	0.42	0.64	0.43	0.41
50-100	0.55	0.78	0.52	0.78
more than 100	1.10	1.70	1.13	1.31

MLE Normed Pain Qs			
label	est	se	P_ratio
male	0.25	0.07	NA
psq	0.26	0.03	NA
intercept	−0.06	0.10	NA
sigma	1.42	0.04	NA
0-25	**1.00**	NA	**1.0000000**
25-50	**2.37**	0.10	**0.4223848**
50-100	**1.87**	0.07	**0.5356272**
more than 100	**0.87**	0.04	**1.1513940**

all questions normed to lowest mean = 1

mean psq control

References

Arrow, K. J. (1950). A difficulty in the concept of social welfare. *Journal of political economy*, *58*(4), 328–346.

Arrow, K. J. (1978). Extended sympathy and the possibility of social choice. *Philosophia*, *7*(2), 223–237.

Arrow, K. J. (2012). *Social choice and individual values* (Vol. 12). Yale university press.

Auspitz, R., & Lieben, R. (1889). *Untersuchungen über die theorie des preises*. Duncker & Humblot.

Aydede, M. (2019). Pain. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Spring 2019 ed.). Metaphysics Research Lab, Stanford University. `https://plato.stanford.edu/archives/spr2019/entries/pain/`.

Bartley, E. J., & Fillingim, R. B. (2013). Sex differences in pain: a brief review of clinical and experimental findings. *British journal of anaesthesia*, *111*(1), 52–58.

Berkouwer, S. B., & Dean, J. T. (2021). Credit, attention, and externalities in the adoption of energy efficient technologies by low-income households. *Unpublished manuscript, University of Pennsylvania, Philadelphia, PA*.

Bishop, R. C., & Heberlein, T. A. (1979). Measuring values of extramarket goods: Are indirect measures biased? *American journal of agricultural economics*, *61*(5), 926–930.

Boyle, K. J. (2017). Contingent valuation in practice. In *A primer on nonmarket valuation* (pp. 83–131). Springer.

Coate, S. (2000). An efficiency approach to the evaluation of policy changes. *The Economic Journal*, *110*(463), 437–455.

Dimand, R. (2019). Irving fisher, ragnar frisch and the elusive quest for measurable utility. Retrieved from `https://ssrn.com/abstract=3495946`

Driver, J. (2014, Sep). *The history of utilitarianism.* Stanford University. Retrieved from `https://plato.stanford.edu/entries/utilitarianism-history/`

Eastmond, T., Mather, N., Ricks, M. D., & Betts, J. (2022). From value added to welfare added: A social planner approach to education policy and statistics.

Finkelstein, A., & Hendren, N. (2020, November). Welfare analysis meets causal inference. *Journal of Economic Perspectives*, *34*(4), 146-67. Retrieved from `https://www.aeaweb.org/articles?id=10.1257/jep.34.4.146` doi: 10.1257/jep.34.4.146

Fisher, I. (1927). *A statistical method for measuring" marginal utility" and testing the justice of a progressive income tax*.

Flood, A. (n.d.). Terry pratchett estate backs jack monroe's idea for 'vimes boots' poverty

index. *The Guardian*. Retrieved 6-27-2022, from https://www.theguardian.com/books/2022/jan/26/terry-pratchett-jack-monroe-vimes-boots-poverty-index

Glick, M. (2018). The unsound theory behind the consumer (and total) welfare goal in antitrust. *The Antitrust Bulletin*, *63*(4), 455–493.

Gorman, W. M. (1955). The intransitivity of certain criteria used in welfare economics. *Oxford Economic Papers*, *7*(1), 25–34.

Hanemann, W. M. (1984). Welfare evaluations in contingent valuation experiments with discrete responses. *American journal of agricultural economics*, *66*(3), 332–341.

Hare, R. M. (1981). *Moral thinking: Its levels, method, and point*. Oxford: Clarendon Press; New York: Oxford University Press.

Harsanyi, J. C. (1955). Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *Journal of political economy*, *63*(4), 309–321.

Harsanyi, J. C. (1986). *Rational behaviour and bargaining equilibrium in games and social situations*. CUP Archive.

Hausman, D. M. (1995). The impossibility of interpersonal utility comparisons. *Mind*, *104*(415), 473–490.

Hausman, D. M., & McPherson, M. S. (2006). *Economic analysis, moral philosophy and public policy*. (Vol. 2nd ed). Cambridge University Press. Retrieved from http://proxy.lib.umich.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=e000xna&AN=149844&site=ehost-live&scope=site

Hendren, N. (2020). Measuring economic efficiency using inverse-optimum weights. *Journal of Public Economics*, *187*, 104198. Retrieved from https://www.sciencedirect.com/science/article/pii/S0047272720300621 doi: https://doi.org/10.1016/j.jpubeco.2020.104198

Hendren, N., & Sprung-Keyser, B. (2020). A unified welfare analysis of government policies. , *135*(3), 1209-1318. Retrieved from https://academic.oup.com/qje/artjournal={QuarterlyJournalofEconomics} doi: 10.1257/jep.34.4.146

Hicks, J. R. (1940). The valuation of the social income. *Economica*, *7*(26), 105–124. Retrieved from http://www.jstor.org/stable/2548691

Jevons, W. S. ([1871] 1911). *The theory of political economy, 4th ed. london: Macmillan*.

Jevons, W. S. (1879). *The theory of political economy*. Macmillan and Company.

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1991). Anomalies: The endowment effect, loss aversion, and status quo bias. *Journal of Economic perspectives*, *5*(1), 193–206.

Kaldor, N. (1939). Welfare propositions of economics and interpersonal comparisons of utility. *The Economic Journal*, *49*(195), 549–552. Retrieved from http://www.jstor

.org/stable/2224835

Kimball, M. S., Ohtake, F., Reck, D., Tsutsui, Y., & Zhang, F. (2015). Diminishing marginal utility revisited. *Available at SSRN 2592935*.

Lautenbacher, S., Peters, J. H., Heesen, M., Scheel, J., & Kunz, M. (2017). Age changes in pain perception: a systematic-review and meta-analysis of age effects on pain and tolerance thresholds. *Neuroscience & Biobehavioral Reviews*, *75*, 104–113.

MacKay, A. F. (1986). Extended sympathy and interpersonal utility comparisons. *The Journal of philosophy*, *83*(6), 305–322.

Mather, N., Eastmond, T., Ricks, M., & Betts, J. (2022). From value added to welfare added: A social planner approach to education policy and statistics.

Miljković, A., Stipčić, A., Braš, M., orević, V., Brajković, L., Hayward, C., ... Polašek, O. (2014). Is experimentally induced pain associated with socioeconomic status? do poor people hurt more? *Medical science monitor: international medical journal of experimental and clinical research*, *20*, 1232.

Morgan, J. N. (1945). Can we measure the marginal utility of money? *Econometrica: Journal of the Econometric Society*, 129–152.

Moscati, I. (2018). *Measuring utility: From the marginal revolution to behavioral economics*. Oxford Studies in History of E.

Pearce, D. (2021, Mar). Individual and social welfare: A bayesian perspectiveâ.

Ruscheweyh, R., Marziniak, M., Stumpenhorst, F., Reinholz, J., & Knecht, S. (2009). Pain sensitivity can be assessed by self-rating: Development and validation of the pain sensitivity questionnaire. *Pain*, *146*(1-2), 65–74.

Ruscheweyh, R., Verneuer, B., Dany, K., Marziniak, M., Wolowski, A., Çolak-Ekici, R., ... others (2012). Validation of the pain sensitivity questionnaire in chronic pain patients. *Pain*, *153*(6), 1210–1218.

Samuelson, P. A. (1950). Evaluation of real national income. *Oxford Economic Papers*, *2*(1), 1–29.

Sen, A. K. (1977). Rational fools: A critique of the behavioral foundations of economic theory. *Philosophy & Public Affairs*, 317–344.

Stephens, R., & Robertson, O. (2020). Swearing as a response to pain: Assessing hypoalgesic effects of novel "swear" words. *Frontiers in Psychology*, *11*. Retrieved from https://www.frontiersin.org/article/10.3389/fpsyg.2020.00723 doi: 10.3389/fpsyg.2020.00723

Stevenson, B., & Wolfers, J. (2019). *Principles of macroeconomics*. Macmillan Higher Education.

Wilson, C. S. (2019). Welfare standards underlying antitrust enforcement: What you mea-

sure is what you get. In *In: Luncheon keynote address at george mason law review 22nd antitrust symposium. arlington: Antitrust at the crossroads.*

Zwolinski, M. (2008). The ethics of price gouging. *Business Ethics Quarterly*, *18*(3), 347–378.