Innovation and Diffusion of Medical Treatment

Barton H. Hamilton
Olin Business School, Washington University in St. Louis

Andrés Hincapié
Department of Economics, Washington University in St. Louis

Robert A. Miller
Tepper School of Business, Carnegie Mellon University

Nicholas W. Papageorge
Department of Economics, Johns Hopkins University

This version: April 24, 2016

Preliminary and Incomplete

Abstract: We construct a dynamic structural model of product choice where product quality is multi-dimensional and where innovation is endogenous to consumer demand. Multiple dimensions of quality imply that new products can be better along some dimensions and worse along others. An example is new medicine that is more effective than existing products, but which also has harsher side effects. In the model, we allow consumer choices, aggregated into market shares, to affect both the speed and the direction of product innovation. The model therefore captures demand externalities since aggregate consumer behavior affects dynamic payoffs through its impact on innovation. We apply the framework to analyze consumer choice and the realized path of innovations over a long time horizon in a maturing product market: HIV drugs. We use the estimated model to assess policies that affect innovation by modifying consumer choices.
Keywords: Innovation, Dynamic Demand, Structural Models, HIV/AIDS.
JEL Classification: O31.

1 Introduction

In many product markets, innovation can lead to substantial quality changes from one point in time to the next. Research on product innovation tends to emphasize demand responses and consumer surplus. Going back to Hicks (1932) economists have recognized that market demand not only responds to, but also drives innovation. Sometimes knows as "demand pull", the idea is that firms respond to consumer preferences by shifting resources towards the development of products that meet potential demand (Schmookler, 1966; Scherer, 1982). Research on demand pull has generally studied how market size or market structure affect the speed of innovation (Finkelstein et al., 2004; Acemoglu and Linn, 2004; Goettler and Gordon, 2011). In contrast, little work has studied demand pull in contexts where product quality is multi-dimensional and where consumer preferences can therefore affect not only the speed, but also the direction of innovation.

In this paper, we introduce an empirical framework to capture how innovation along multiple dimensions of quality is endogenous to aggregate consumer demand. Allowing multiple dimensions of product quality means that new products can be better on one dimension and worse along another. An example is new medicines that are more effective at curing illness, but have harsher side effects. Our framework centers around estimating the multi-dimensional empirical distribution of innovations, which is then embedded into a structural model of dynamic demand. In the model, forward-looking consumers make choices after forming expectations over potential future innovations. Optimal consumer choices are then aggregated into market shares, which help to drive both the speed and the direction of innovation by determining how new products are drawn from the empirical innovation distribution. In this sense, the model takes explicit account of demand externalities, which arise since aggregate consumer behavior affects dynamic payoffs through its impact on innovation. We match our model to data on the realized path of innovations, product quality and consumer choices over a long time horizon in a maturing product market: HIV drugs.

Our framework captures several layers of consumer uncertainty. First, consumers form expectations over the qualities of the products available to them and become fully aware of their qualities only after they have used a product at least once. Alternatively, consumers may experiment with new technologies that are not yet on the market and may mark a substantial improvement over existing products. But experimental technologies may also be of dangerously low quality. In software, this is known as *beta-testing*; in medicine, this is done through participation in clinical trials. Second, consumers make decisions to maximize their lifetime utility, accounting for how current-period choices affect their future utility. Though not an important feature in some product markets, dynamics are crucial to under-

standing choices in many instances, e.g., when goods are storable or have a lasting impact on utility, as is the case with pharmaceuticals or medical treatments. Third, in a market where substantial innovation is possible, consumers must form expectations over the path of technology, effectively forecasting how product markets will evolve.

A key contribution of our framework is the manner in which we model consumers expectations over market evolution, in particular, the number and qualities of potential new technologies. The observed path of product innovation is explicitly modeled as a single draw from an underlying stochastic process. Consumers choosing among existing products are tasked with forming expectations over this stochastic process, fully aware that aggregate demand ultimately drives the path of innovation. In forming expectations, each consumer also takes explicit account of how every other consumer likewise forecasts technological advancements when making decisions. Notice, each consumer's choice behavior is therefore a function of all consumers' beliefs about future product markets.

By modeling innovation as a function of demand, we can assess how different policies, through their impact on demand, can also influence the path of technological innovation. For example, if we encourage consumers who are reluctant to try new products with uncertain qualities, can we increase the likelihood of breakthroughs? Moreover, given demand externalities, it is possible that a planner could solve the consumer coordination problem wherein patients underuse products that could raise dynamic payoffs through their impact on innovation.

We apply our framework to the market for HIV drugs. HIV is a medical condition that reduces the ability of the immune system to fight off routine infections (a condition known as AIDS).[1] It reached epidemic proportions in several countries, including the U.S., starting in 1984. HIV has reached a point where—at least in developed countries where access to medication is widespread and subsidized—the condition is manageable and side effects of medications are fairly mild. However, this was not always the case. In the early years of the epidemic, available treatments were not only largely ineffective, but also had uncomfortable, painful and even deadly side effects. Each year brought innovations that were incremental at best. Indeed, as we will show, some new products were worse since they were more toxic without being more effective. In the mid-nineties, a new set of treatments (collectively known as HAART) was introduced, which effectively transformed HIV from a virtual death sentence to a chronic condition.[2] Within two years, the introduction of HAART reduced mortality

[1] AIDS stands for acquired immunodeficiency syndrome.

[2] HAART stands for highly active anti-retroviral treatment. There is no vaccine or cure for HIV or AIDS, but HAART is the current standard treatment. In general, 1996 is marked as the year when two crucial clinical guidelines that comprise HAART came to be commonly acknowledged. First, protease inhibitors (made widely available towards the end of 1995) would be an effective HIV treatment. Second, several

rates by over 80% among HIV+ men (Bhaskaran et al., 2008). HAART therefore marked a clear departure from existing products in the market for HIV treatments. However, HAART involved drugs that were highly toxic, leading to side effects that were often intolerable and drove some people to avoid using them. In other words, HAART comprised treatments that were better on some dimensions, but worse on others. Thereafter, a series of new drugs were introduced, which were effective and had fewer side effects.

We use data on HIV+ men's treatment decisions and health outcomes over approximately 20 years. The benefit of observing a long panel in the market for HIV drugs is that we can watch the path of innovation unfold. Indeed, a number of breakthroughs have occurred since the early, darker years of the AIDS epidemic. Since, we observe the same individuals over time, the evolving market allows us to identify both endogenous innovation and consumer preferences. In the model, consumers form expectations about the process underlying this path. In our framework, we exploit the fact that we can observe consumer decisions over time and the realized path of innovation to better understand how expectations were formed. *Ex post*, we can observe that the path of technological innovation occurred in fits and spurts. The introduction of HAART constitutes a key source of variation, which we exploit to help identify our model. Indeed, the realized path of innovation includes fairly incremental changes to drug qualities along with massive innovations that drastically altered the lives of consumers with HIV. In our framework, large and drastic changes in the product market are draws (albeit less likely ones) from the same underlying distribution that generates smaller, incremental improvements.

We contribute to three separate literatures. First we contribute to literature studying how consumer behavior affects innovation. Schmookler (1966) formalized the idea, calling it "demand pull". Building on this idea, several papers have demonstrated that market size affects the speed of innovation. For example, Finkelstein et al. (2004) show that policies promoting vaccine use accelerate the development of vaccines. Related, Dranove et al. (2014) identify a "social value" of pharmaceutical innovation, showing that Medicare Part D spurred the development of some drugs. If consumer behavior drives innovation which benefits other consumers, it follows that a demand externality arises.[3] In the context of obesity, Bhattacharya and Packalen (2012) provide evidence that individual efforts to prevent obesity can shrink the market size for obesity treatments, which slows technological progress. If so, individuals may over-invest in preventative care compared to the social optimum. A

anti-retroviral drugs taken simultaneously could indefinitely delay the onset of AIDS.

[3]Demand externalities have been discussed in a variety of scenarios. For example, Allcott et al. (2015) speculate that consumers who prefer unhealthy food may exert a negative externality on healthy eaters by discouraging the purveyors of healthy foods from entering a market, thus contributing to the emergence of "food deserts".

similar idea applies to experimentation with new products. Bolton and Harris (1999) argue that a free-riding problem emerges if experimenting accelerates innovation. This relates to our context if clinical trials provide social benefits by spurring innovation, in which case individually rational consumers may choose to participate less than is socially optimal.

Related to the idea of "demand pull", Goettler and Gordon (2011) show that market structure also drives innovation. They find that in the market for computer processors, the presence of a second firm can slow innovation (since firms do not expect to capture all profits), but that consumer surplus falls due to monopolistic prices. In contrast to Goettler and Gordon (2011), we add a second dimension to product quality in a model where innovation is endogenous to demand, which means that consumer preferences can affect the direction and the speed of innovation.[4] In a setting where product quality is multi-dimensional, consumers may benefit from innovations along one dimension of quality, but prefer to use drugs that would encourage innovation along another dimension. For example, consumers may be reluctant to experiment with new drugs that are not very effective, but have few side effects, even though doing so could potentially encourage the development of drugs that are highly effective and also have fewer side effects than existing drugs.

A second literature we contribute to studies dynamic demand under uncertainty. Following Petrin (2004), each product in our model is a bundle of characteristics.[5] Moreover, in our framework, characteristics can have dynamic impacts on consumers (Gowrisankaran and Rysman, 2012). Literature on product choice has considered the idea that consumers are unaware of product characteristics or match value. Erdem and Keane (1996) study the value of experimentation with new products to learn about their qualities. Learning has been incorporated into dynamic models of pharmaceutical demand. Examples are Crawford and Shum (2005) and Chan and Hamilton (2006), where the latter paper explicitly discusses the role of side effects. We incorporate learning and uncertainty into our model in several ways. First, and similar to existing work, we model consumers as learning about existing market products that they have never used. Second, consumers can experiment with new products that are not yet widely available by participating in a clinical trial. Third, we depart from existing work on dynamic demand in how we model consumer expectations over the path of innovation. Most papers take the existing set of products as given or exogenous to the model and focus on demand responses to new products. In contrast, we explicitly

[4]It is important to point out that, unlike Goettler and Gordon (2011), we do not explicitly model firm interaction or dynamic decisions. Therefore, we are unable to conduct policy analysis related to market structure using our framework. An interesting extension of the current paper would be to merge the two approaches by integrating firm decision-making into a model where products have multiple qualities.

[5]Studies pioneering the 'characteristics approach' include Stigler (1945), Lancaster (1966) and Rosen (1974).

model how consumers form expectations about future innovations, and allow them take into account that aggregate market shares can shift the direction of innovation.

Methodologically, we build on Hotz and Miller (1993) and Hotz et al. (1994) in using forward simulation to incorporate how individuals form expectations about future innovations. In our context, the choice set that individuals face is non-stationary. We handle this problem by re-defining the current state of technology using a stationary distribution of innovations and a non-stationary reference point or centroid for innovation that emerges endogenously from consumer demand. This is similar to what Goettler and Gordon (2011) do in their framework when studying microprocessor speed. However, in their setting, product quality is one-dimensional and the innovation distribution is effectively binary (either improving by a fixed amount or not). In our case, we need to account for demand externalities where product quality is multi-dimensional, which means that new product qualities can move in many different directions on a two-dimensional plane. Moreover, as we show, the empirical distribution of innovations for HIV drugs is not well-approximated as movements with a fixed distance. In light of these features of our setting, when computing lifetime utility associated with each choice, we use forward simulation to capture how consumers make decisions after forming expectations about potential future innovations.

The remainder of this paper is organized as follows. Section 2 describes the data set we use. In Section 3, we specify the structural model and in Section 4 we discuss estimation. In Section 5, we present parameter estimates and describe model implications for the distribution of innovations. In Section 6, we use the estimated model to conduct counterfactual policy simulations. Section 7 concludes.

2 Data

In this section we introduce the data set used in this paper and describe some of the key empirical patterns we use to identify structural parameters. We use the public data set from the Multi-Center AIDS cohort Study (MACS). The MACS is an ongoing longitudinal investigation (beginning in 1984) of HIV infection in men who have sex with men (MSM) conducted at four sites: Baltimore, Chicago, Pittsburgh and Los Angeles.[6] At each semi-

[6]Data in this manuscript were collected by the Multi-Center AIDS Cohort Study (MACS) with centers (Principal Investigators) at The Johns Hopkins Bloomberg School of Public Health (Joseph B. Margolick, Lisa P. Jacobson), Howard Brown Health Center, Feinberg School of Medicine, Northwestern University, and Cook County Bureau of Health Services (John P. Phair, Steven M. Wolinsky), University of California, Los Angeles (Roger Detels), and University of Pittsburgh (Charles R. Rinaldo). The MACS is funded by the National Institute of Allergy and Infectious Diseases, with additional supplemental funding from the National Cancer Institute. UO1-AI-35042, 5-MO1-RR-00052 (GCRC), UO1-AI-35043, UO1-AI-35039,

annual visit, survey data are collected on HIV+ men's treatment decisions, out-of-pocket treatment expenditures, physical ailments, which can reflect drug side effects, along with sociodemographic information, such as labor supply, income, race, and education.

In addition, blood tests are administered at each visit to objectively measure health status. Our main objective measure of immune system health is *CD4 count*, defined as the number of white blood cells per cubic millimeter of blood. Absent HIV infection, a normal range is between 500 and 1500. For HIV+ individuals, a count below 500 indicates that the immune system has begun to deteriorate due to HIV, but can still fight off infections such that the individual is not symptomatic. When CD4 count drops below about 300, a patient is said to suffer from AIDS.[7] AIDS means that the immune system becomes unable to fight off routine infections and survival probability drops.

2.1 Summary Statistics

The full MACS data set contains information on 6,972 subjects at 49 possible semi-annual visits for a total of 111,271 observations in the form of subject-visit dyads. We limit our attention to HIV+ individuals, leaving us with 47,753 observations. Due to lack of data on gross income and out-of-pocket treatment costs at earlier visits, we drop observations prior to visit 14 (roughly, late 1990) and for robustness in the reporting of survival we also drop observations after visit 47 (about 2008). These sample period restrictions leave us with 29,523 observations and 2,420 individuals. Next, we drop observations where data are missing on at least one of the variables used in subsequent analysis (though we conduct various robustness checks to insure that our results are not driven by these exclusions). After these exclusions, the remaining analytic sample consists of 1,719 unique individuals and 16,851 observations.

Summary statistics by individual are reported in Table 1. The first column presents statistics for the analytic sample.[8] 68% of sample subjects are white, 22% are black and about 9% are hispanic. Race variation is important since previous research has emphasized difficulties in recruiting blacks into clinical trials, which may reflect different costs associated with treatments or variation in expected health outcomes. About 86% of the sample received some secondary education or more and nearly a quarter (23%) attended graduate school. Consistent with previous research studying medication choice using the MACS data set, there is evidence of substantial variation in labor supply (Papageorge, 2016). 74% of the

UO1-AI-35040, UO1-AI-35041. Website located at http://www.statepi.jhsph.edu/macs/macs.html.

[7]AIDS stands for acquired immunodeficiency syndrome. The CD4 cutoff below which AIDS occurs varies between 200 and 350.

[8]For comparison, the third column reports statistics for a larger sample of 2,420 individuals, where we have not dropped observations due to missing data on any particular variable.

sample is observed working at least once and 68% of the sample observed not working at least once.

Underscoring the seriousness of HIV infection, about 40% of the HIV+ subjects we observe at least once over the sample period die prior to the end of the sample period. However, product market innovation led to drastic changes for HIV+ men. The most striking example is the introduction of HAART in the mid-1990s, which was much more effective at improving underlying health compared to the treatments that preceded it. Conditional on surviving until the invention of HAART, 20% of subjects are observed dying. This understates the impact of HAART since the sample under study is an aging cohort, i.e., survival rates are much higher when the cohort is considerably older when HAART is available. Further, according to Table 1, about 83% of subjects are observed using a market product at least once. Moreover, nearly a quarter (24%) opt for early access by participating in a clinical trial at least once during the sample period, suggesting that patients are willing to try experimental products where quality is uncertain.

2.2 Innovation and Consumer Demand

In this section, we consider pre versus post-HAART behavior to investigate how consumers respond to pharmaceutical innovation. Since HAART marked a large innovation on earlier treatments, it lead to strong and observable consumer responses that help identify consumer preferences. Summary statistics for subject-visit dyads are found in Table 2 for the full analytic sample (column [1]) and then separately for the pre and the post-HAART eras (columns [2] and [3], respectively). We split the sample by HAART era to illustrate substantial changes to choices and outcomes after HAART was introduced.

Perhaps the most striking example of the impact of HAART on consumers is through its effect on survival. In Figure 1(a), we plot the probability of dying between periods t and $t+1$ conditional on survival until t. Death rates are much higher prior to HAART introduction and despite a multitude of new treatments coming available. After HAART, death rates plunge, and continue to fall until 2007, as smaller innovations occur that make drugs incrementally more effective and less toxic. In Figure 1(b), we plot average CD4 count over time for people on market drugs and no treatment for HIV. Over time, health for people taking no drugs remains fairly constant while health for individuals in a market drug rises.

Notice that average age rises and labor supply and income decline after HAART, consistent with the fact that we observe an aging cohort, which is more likely to retire and report lower gross income over time. HAART introduction also affected immune system health, as measured by CD4 count. According to Table 2, average CD4 count among HIV+ men

in our sample is 407 in the pre-HAART era, rising to 524 in the post-HAART era. Interestingly, after HAART, the proportion of individuals reporting physical ailments declines only slightly (45% to 41%). The small change reflects the net effect of two countervailing dynamics (Papageorge, 2016). HAART improved health on average, which lowered reported ailments attributable to symptoms of HIV. However, HAART also led to side effects among users, thereby increasing reports of ailments. The increase in side effects also reflects how use of HIV treatment rose with the introduction of HAART, from 45% to 76%.

The rise in consumption of HIV treatments after HAART was introduced suggests that patients are more likely to use drugs despite side effects if the utility cost of suffering ailments is offset by expected improvements to health. HAART was more effective than earlier drugs, which encouraged people to use it despite its side effects. Another shift in consumption of HIV treatments occurred after HAART and in relation to participation in clinical trials. Trial participation in the post-HAART era is about half its pre-HAART level. This drop is consistent with the idea that individuals are willing to experiment with potentially low-quality technologies by trying new drugs if no good market products are available. However, once innovation has led to technological advances, individuals are less willing to experiment with potentially low quality drugs.

Given the impact of HAART on health, it is important to understand why many consumers did not use it. In Figure 1(c), we plot the proportion of HIV+ consumers using an HIV treatment. Notice that treatment consumption is about 50% in 1990 and actually falls prior to HAART introduction. This reflects that products available on the market are of fairly low quality. Still, if quality were uni-dimensional, even a low quality drug would be better than no drug at all. Treatment costs are one possibility. In Table 2, we see that treatment costs rise after HAART introduction, from about \$179 to \$327 for six months of treatment. In Figure 1(d), we plot out-of-pocket treatment costs over time for individuals who use drugs (takers) and for individuals who are not on any HIV drugs (non-takers). We see that costs are low (averaging between 300\$ and 600\$ per year for individuals earning on average about \$37,000 per year). Second, non-users of market drugs pay non-zero costs for drugs, perhaps spending more money on medication to fight opportunistic infections. In other words, the incremental out-of-pocket cost of HIV treatments does not appear sufficient to explain why some people avoid HIV treatments.

Another possibility is that drug quality is multi-dimensional in which case demand reflects a distaste for another feature of HIV drugs. Given data on physical ailments, we explore the possibility that consumer demand reveals a distaste for side effects. This would explain the rapid rise in use of HIV treatments after HAART is introduced since individuals would be more willing to use drugs with side effects as long as drugs are effective at improving

underlying health. Evidence in support of this idea is market consumption by CD4 count, plotted in Figure 2(a). Sicker people are far more willing to take low effective medications despite side effects in the years before HAART. After HAART, notice a striking convergence in the proportion of men using medications, driven largely by healthy individuals going onto medication. We also plot physical ailments over time in Figure 2(b). For non-users of HIV medications, ailments remain fairly steady. For users of HIV medications, ailments drop prior to HAART introduction and then rise after HAART, which is consistent with HAART being a highly effective drug with side effects. However, after 2001, ailments decline for individuals using HIV drugs. This reflects later improvements to medications, which lowered their side effects.

Another option for individuals in the product market we study is to join a clinical trial to gain early access to new products. Studying how individuals experiment with new drugs by joining a clinical trial further highlights how consumers respond to innovations in the market for HIV drugs. Trial participation over time is plotted in Figure 2(c). Trial participation hovers around 4%, but begins to rise in the years leading up to HAART introduction. This reflects two dynamics. First, as individuals became ill, they were more willing to experiment with new products of uncertain qualities. Second, in the years just prior to HAART introduction, the drugs that comprise HAART, including protease inhibitors, marked a substantial improvement over drugs available on the market. In those years, trial participation gave individuals early access to much better products. This relates to the idea of *beta testing* in markets where some consumers are willing to experiment with new products with high potential quality. Finally, notice that trial participation plunges after HAART is introduced as a market option. The reason is that individuals no longer need to participate in a clinical trial (and face therefore more uncertainty) to access good drugs. In support of this idea, we consider participation by CD4 count in Figure 2(d). Notice, early trial participation is driven largely by individuals with low CD4 counts. After HAART, there is a marked convergence, which means that once effective drugs are available, it is no longer possible to explain trial participation as an option for people who are very sick and therefore willing to face uncertainty in exchange for early access to a high-quality product.

2.3 Market-Level Innovation

In the previous section, we studied how consumer responses to innovation shed light on consumer preferences. The patterns we have described until now are consistent with the idea that patients value their health, but are also concerned with side effects. Moreover, side effects seem to play a larger role in demand after survival is more or less assured. However,

our preliminary evidence also suggests that preferences are not lexicographic. Patients seem willing to use toxic (or experimental) medication if the alternative is a large rise in the probability of dying, but patients will also forgo treatments with harsh side effects if drugs are not effective and the survival gains are limited.

In this section, we consider market-level innovation. To start, we illustrate innovation and diffusion of new products over time in the market for HIV treatment using a "heat map" displayed in Figure 3. For the approximately 90 drugs that were most used, we compute market share over our sample period.[9] Dark blue corresponds to no market share and warmer colors mean higher market shares. In the early years of the epidemic, there are only a few drugs with high market share. Over time, many new drugs emerge, most with lower market share. The heat map captures two important patterns in the data. First, many new drugs were invented over time. In other words, the market for HIV drugs was active over our sample period. Second, most old drugs eventually exit and are replaced by new drugs, which means that new drugs marked improvements upon older ones. A striking shift occurs in the mid-1990's, after which point most earlier drugs exit, replaced by new drugs. This corresponds to HAART introduction, when protease inhibitors (PI's) were introduced and became a standard part of HIV treatment. After HAART, moreover, many drugs became obsolete.[10]

Finally, we ask whether the observed innovation path can be seen as a response to consumer preferences. In other words, we search for evidence of the idea that demand drives the direction of innovation. In Figure 4, we plot drug qualities (effectiveness and side effects) for different periods of time. The figure illustrates the path of technology over time. Notice how there is a large innovation in the direction of improved health in the mid-1990's. This improvement is the introduction of HAART. Moreover, notice that there is some evidence of a rightward shift in later years as innovations reduce side effects without offering much of an improvement in efficacy. This rightward shift is important as it corresponds to changes on the relative importance of one dimension of taste over another. We argued previously that consumer demand patterns, as the market for HIV drugs matured, seem to show a preference for drugs with fewer side effects, especially when survival is less of a concern. The path of innovation seems to have followed shifts in market demand after HAART was introduced. Therefore, preliminary empirical patterns provide support for the idea that

[9] Appendix A is a data appendix that contains additional information on individual drugs and treatment combinations. Table S1 discusses which drugs or combinations are taken in clinical trials. Table S2 lists the chemical compositions of each drug. Table S3 shows how drugs are combined into treatments. Table S4 discusses "core treatments", which are the main sets of treatments we observe, including the individual drugs they are composed of, whether or not they count as HAART and their entry and exit visits.

[10] An exception is AZT, which remained a standard component of HAART.

innovation responded to consumer demand.

3 Model

We specify a model in which individuals maximize lifetime utility by choosing an HIV medical treatment. Medical treatments can improve health and increase lifespan, but have potential side effects, which affect survival and labor market outcomes. Individuals can choose an HIV treatment that is available on the market, opt for no treatment at all or experiment with a new treatment by participating in a clinical trial.

In making decisions, individuals face several sources of uncertainty. First, they are uncertain about current-period outcomes, including their income and side effects, all of which are affected by their treatment decision. Second, individuals are uncertain about the evolution of other individual-specific state variables, notably health and survival, which are likewise affected by medical treatment. Finally, individuals face uncertainty over the evolution of the product market since new treatments may enter the market and some incumbent treatments may drop out.

In specifying the model, we begin with the individual's problem. We introduce state variables, choices, flow utility and the stochastic processes governing current-period outcomes and state-to-state transitions. Using these components we specify the value function. Next, we discuss the evolution of the aggregate state, which will be endogenous to individual choices. In particular, we discuss entry of new treatments and exit of incumbent treatments.

3.1 The Individual's Problem

The individual chooses medical treatment to maximize expected discounted lifetime utility. In making decisions, he observes his current set of state variables which includes individual-specific variables, such as health, and market-level variables, such as the state of medical technology. Individuals use market-level variables to form expectations over the future path of innovation. In specifying the individual's problem, we begin by introducing state variables, choice set, flow utility and stochastic processes governing outcomes and state-to-state transition probabilities. Next, we specify the value function.

3.1.1 State Variables

The state for individual i at period t is denoted $\mathcal{Z}_{it}$, where

$$\mathcal{Z}_{it} \equiv \langle z_{it}, \varepsilon_{it} \rangle \tag{1}$$

z_{it} is a set of state variables that is further sub-divided into a set of individual-specific variables, denoted $z_{it}^{\mathcal{I}}$, and a set of aggregate variables denoted $z_t^{\mathcal{M}}$:

$$z_{it} \equiv \left\langle z_{it}^{\mathcal{I}}, z_t^{\mathcal{M}} \right\rangle \tag{2}$$

The individual-specific state variables, $z_{it}^{\mathcal{I}}$, are

$$\begin{aligned}
b_i &: \text{a set of race indicators} \\
edu_i &: \text{a set of education indicators} \\
h_{it-1} \in \mathbb{R}_+ &: \text{health at the start of period } t \\
a_{it-1} \in \{25, 25.5, \ldots\} &: \text{age at the start of period } t \\
l_{it} \in \{0, 1\} &: \text{working during period } t \\
q_{it-1} = \left\{q_{it-1}^x, q_{it-1}^h\right\} \in \mathbb{R}^2 &: \text{characteristics of product consumed last period} \\
\eta_i &: \text{person-specific income characteristic}
\end{aligned}$$

The individual can be either white, black or Hispanic. He belongs to one of four mutually exclusive educational categories: high school, some college, college, and more than college. His health, measured by CD4 count, is a continuous positive number.[11] His age is measured in half-year increments, corresponding to the frequency of MACS data collection. l_{it} indicates whether he will work during period t.

Each HIV treatment has two characteristics: its effectiveness at raising CD4 count, which we denote θ^h, and its propensity to cause side effects, denoted θ^x. We collect these into a vector denoted $\theta \in \mathbb{R}^2$. If the individual consumed a market product in the prior period, the characteristics of that product, denoted q_{it-1}, are part of his current state space. Finally, all elements of $z_{it}^{\mathcal{I}}$ are observed to the econometrician except η_i, which is an exogenous person-specific characteristic that affects the income process and is described below.

Besides individual-specific variables, z_{it} contains variables summarizing the current aggregate landscape defined as

$$z_t^{\mathcal{M}} \equiv \langle W_t, \omega_t, \mathcal{F}_t \rangle, \tag{3}$$

[11] CD4 ranges from 0 to 2915 in our analytic sample with a median of 448. Healthy CD4 counts are those above 500 units per mm^3 and typically range between 500 and 1,500.

Individuals use $z_t^{\mathcal{M}}$ to forecast the evolution of the market. W_t consists of characteristics of market products that are available at time t. ω_t is a point in the product space that summarizes the current state of technology. Individuals use ω_t to make forecasts about new drugs that may enter the market in the future. $\mathcal{F}_t$ is the aggregate joint distribution of consumer characteristics which individuals use to compute expectations over aggregate behavior determining the future evolution of the market. The elements of $z_t^{\mathcal{M}}$ will be described in further detail below, when we discuss how we model the evolution of the product market.

Finally, individuals also face a vector of choice-specific additive utility disturbances ε_{it}, which are unobserved to the econometrician and assumed independent across time, individuals and choices.

3.1.2 Choices

At each period t individual i chooses whether or not to use medication. If he opts for medication, he may choose the same product he consumed in the last period or he may choose from the set of other treatments that are currently available on the market. Alternatively, he may choose a trial treatment. The individual faces uncertainty about the quality of both market and trial treatments.

We begin with uncertainty over market treatments. If the individual chooses the same market treatment he consumed in the prior period, he faces no uncertainty regarding its characteristics. As discussed in the previous section, the individual's state space includes the characteristics of the drug consumed in the prior period q_{it-1}. On the other hand, if the individual chooses a different market drug, his alternative is to choose one among several groups or *clusters* of drugs with similar qualities. The agent is then randomly assigned a drug within the cluster he selected.

Formally, at every period t there is a set of market products $\mathcal{P}_t$ clustered in several groups collected in $\mathcal{G}_t$. $\mathcal{G}_t$ denotes both the collection of clusters available at t and the cardinality of the collection. When individual i decides to consume a market treatment that is different from the one he consumed in the prior period, he must choose from a cluster $g_t \in \mathcal{G}_t$. By selecting group g_t he chooses a gamble among all products in group g_t. The distribution of products within the group is given by weights that are a function of the treatment characteristics and the number of products in the group. The estimation of these weights is explained below. Even though the moments of the within cluster distribution are generated by the underlying products in the cluster and their weights, we assume that agents only observe the first two moments of the within cluster distributions and that these are sufficient to describe the distribution.

Our clustering process is a device to make estimation feasible by reducing the state space significantly while still allowing individuals to choose among different options in the market.[12] In order to avoid scaling issues when using our clustering algorithm, we assume that clustering occurs with respect to scaled product characteristics denoted $\tilde{\theta} \in [-1,1]^2$ that are explained in the estimation appendix (Appendix B). Then we obtain product groups at t by solving a k-means algorithm that approximates the solution of the following objective function[13]

$$\min_{1\{k\in g\}_{k\in\mathcal{P}_t}|\mathcal{G}_t} \quad \sum_{g=1}^{\mathcal{G}_t}\sum_{k\in\mathcal{P}_t} 1\{k\in g\}\left\|\tilde{\theta}_k-\tilde{\theta}_k^c\right\|^2$$
$$s.t.\sum_{g\in\mathcal{G}_t} 1\{k\in g\}=1 \text{ for all } k\in\mathcal{P}_t \tag{4}$$

where the centroid of cluster k, $\tilde{\theta}_k^c$, is defined as

$$\tilde{\theta}_k^c=\frac{\sum_{k\in\mathcal{P}_t} 1\{k\in g\}\tilde{\theta}_k}{\sum_{k\in\mathcal{P}_t} 1\{k\in g\}} \tag{5}$$

The algorithm is explained in detail in Appendix B. At any given period we set the maximum value of $\mathcal{G}_t$ at $\mathcal{G}^{\max}$ so that the individual knows how many groups will be available every period. $\mathcal{G}^{\max}$ is chosen so that there is a non negligible number of consumers choosing each group in the data. We set $\mathcal{G}^{\max}=3$.

We do not model the variation of within cluster assignment endogenously. Instead, we develop the concept of within cluster weights as functions of products's characteristics. Weights are estimated in the following fashion:

1. We compute a nonlinear regression of within cluster shares on treatment characteristics:

$$s_{k|g_t}=\exp\left(X_{k,t}^w\beta^w\right)+\epsilon_{k|g_t}^w \tag{6}$$

   where $X_{k,t}^w$ includes a constant term, the ranking (within its cluster) of the characteristics of the product, the number of members in the cluster, whether the product is new, and several interactions.

2. We obtain predicted within cluster shares $\hat{s}_{k|g_t}$ and compute the weight of product k

[12] This approach is close to reality if individuals only observe product labels and do not know their characteristics beyond the fact that groups of product labels are associated to a certain mean and variance of characteristics.

[13] See Duda and Hart (1973) and Andrew W. Moore's *K-means and Hierarchical Clustering* tutorial at http://www.cs.cmu.edu/∼awm/tutorials.html.

in cluster g_t as

$$\tilde{s}_{k|g_t} = \frac{\hat{s}_{k|g_t}}{\sum_{r \in g_t} \hat{s}_{r|g_t}} \tag{7}$$

If the individual chooses neither to try a cluster nor to stay in his previous treatment, he may instead experiment with a new drug that is available only in a clinical trial. Trial product characteristics are unknown, but are distributed according to $F_{\theta|\omega_t}$, where ω_t, discussed below, characterizes the distribution from which experimental drugs will be drawn. A key difference between consuming group g_t and the trial treatment is that once the individual chooses a group a treatment is assigned to him, he has the chance of choosing that treatment with certainty the next period.

Having described each option, we now formally specify the choice set. Let d_{jit} be the choice indicator that takes the value of one if agent i in period t chooses medical treatment j in the choice set $\mathcal{C}_{it}$. Notice, the choice set is time-specific since the market for available products evolves as new products enter the market and incumbent products exit. The choice set is also individual specific since individuals who chose a market treatment in the prior period may choose that treatment again. If the individual did not choose a market treatment in the prior period his choice set is:

$$\mathcal{C}_{it} = \begin{cases} 0 & \text{No Treatment} \\ 1 & \text{Cluster } g_t = 1 \\ 2 & \text{Cluster } g_t = 2 \\ \vdots & \vdots \\ \mathcal{G}^{max} & \text{Cluster } g_t = \mathcal{G}^{max} \\ \mathcal{G}^{max} + 1 & \text{Trial} \end{cases} \tag{8}$$

If the individual chose a market treatment in the prior period his choice set $\mathcal{C}_{it}$ is augmented by one alternative to include the possibility of consuming his previous period treatment again.

3.1.3 Utility

Next, we specify the flow utility function to capture how the individual's product choices are driven by the effects of each treatment choice on health, ailments, income, out-of-pocket payments, and non pecuniary benefits. For choice $j \in \mathcal{C}_{it}$ and state z_{it}, the utility at period t for individual i is a function of his health, ailments, and net income given by

$$y_{jit} + \varepsilon_{jit} = \alpha_{jit}(z_{it}) + \alpha_m(m_{jit} - o_{jit}) + \alpha_x x_{jit} + \alpha_{xp} x_{jit}(1 - d_{0it}) + \varepsilon_{jit} \tag{9}$$

where m_{jit} is gross income, o_{jit} are out-of-pocket payments, x_{jit} is an indicator for whether the individual does not suffers ailments, d_{0it} is the indicator of whether he chooses not to consume a treatment, and ε_{jit} are unobserved choice-specific taste shocks that are independent over time as well as across alternatives and individuals. The interaction of the no-ailments indicator and the treatment choice indicator is used to capture a distaste for side effects, which are ailments arising from treatment consumption.

In equation (9), $\alpha_{jit}(z_{it})$ are choice-specific preference parameters that depend on observables. They are defined as

$$\alpha_{jit}(z_{it}) \equiv \alpha'_{jb} b_i + \alpha_{ja} a_{it-1} + \alpha_{jh} h_{it-1} \tag{10}$$

We assume that consumer preferences over clusters are fully captured by cluster characteristics. Therefore, we assume parameters α'_{jb}, α_{ja}, and α_{jh} to be constant across clusters. This is the characteristics approach commonly used in structural models of demand which explains consumer choices as a function of product qualities. In contrast, participating in a clinical trial may offer differential benefits related to the psychological costs (or benefits) from being part of an experiment. We also allow the choice of remaining in the same product to have differential non pecuniary benefits in order to capture factors, such as switching costs, which explain why consumers may continue using a product they have used before even as better products enter the market. Finally, we normalize the non pecuniary benefits from not consuming a treatment to zero.

3.1.4 Outcomes and Transitions

In this section, we specify the stochastic processes governing state variables in z_{it} as well as the outcome variables: income, out-of-pocket payments, ailments, and survival.

Income: Gross income is a function of today's state, z_{it}, and ailments, x_{jit}. It is given by

$$m_{jit} = X^m_{jit}\Gamma^m + \eta_i + \epsilon^m_{it} \tag{11}$$

where $X^m_{jit} = [1, h_{it-1}, \ldots, h^7_{it-1}, a_{it-1}, a^2_{it-1}, b_i, edu_i, l_{it}, x_{jit}]$. Gross income does not include product cost, which is accounted for in the payments equation below. Equation (11) is estimated using random effects and individual-specific income characteristics are estimated consistently as

$$\hat{\eta}_i = \sum_t \sum_j d_{jit}\left(m_{jit} - X^m_{jit}\hat{\Gamma}^m\right)$$

Individuals observe the income iid shocks ϵ^m_{it} before making their choice.

Payments: Out-of-pocket payments are censored at zero. They are given by the following

tobit specification

$$o_{jit} = o\left(X^o_{jit}, \epsilon^o_{it}; \Gamma^o\right) \tag{12}$$

where $X^o_{jit} = [1, h_{it-1}, \ldots, h^6_{it-1}, a_{it-1}, a^2_{it-1}, b_i, edu_i, \{d_{jit}\}^5_{j=0}, l_{it}, x_{jit}]$ and ϵ^o_{it} is the error term in the underlying equation. Since we do not directly observe prices, and in order to simplify the problem, we assume a constant cost of participating in a trial as well as a constant cost of consuming a market product.[14]

Labor Supply: We do not model labor supply explicitly as a choice as it is not the main purpose of this paper. However, labor supply may be affected by treatment choices, e.g., through health status and physical aliments. Moreover, labor supply also affects income and therefore utility. To capture this, we treat labor supply as a state variable that individuals know at the beginning of the period before making their treatment decision. Individuals draw their labor market participation from the distribution characterized by

$$\Pr[l_{it} = 1|X^l_{it}] = \frac{1}{1 + \exp(X^l_{it}\Gamma^l)} \tag{13}$$

where $X^l_{it} = [1, l_{it-1}, h_{it-1}, \ldots, h^4_{it-1}, a_{it-1}, a^2_{it-1}, b_i, edu_i]$.

Physical Ailments: First, define the characteristics of the treatment as a function of the choice as

$$\theta(d_{jit}) = \{\theta^x(d_{jit}), \theta^h(d_{jit})\} \tag{14}$$

where $\theta(d_{jit}) = q_{it-1}$ if the individual consumes his prior-period market treatment.

Ailments are determined by a production function the inputs of which are drug characteristics and health. Let x_{jit} be an indicator that takes the value of 1 if the individual does not suffer ailments in t. The probability of not having physical ailments for individual i choosing $j \in \mathcal{C}_{it}$ at time t is modeled as:

$$\Pr\left[x_{jit} = 1|\cdot\right] = \frac{\exp\left(\sum^5_{m=0} \gamma^x_m h^m_{it-1} + \theta^x\left(d_{jit}\right)\right)}{1 + \exp\left(\cdot\right)} \tag{15}$$

Health: CD4 count is our objective measure of health. Like ailments, health at the beginning of period $t+1$ is a function of drug characteristics and health. The health production function is specified as:

$$h_{jit} = \sum_{m=0}^{5} \gamma^h_m h^m_{it-1} + \theta^h\left(d_{jit}\right) + \epsilon^h_{it} \tag{16}$$

[14]End-users customarily pay a standardized deductible that is a fraction of the brochure price of the drug paid by the insurance company. Median out-of-pocket drug costs are about $300 every six months for a regime of drugs that would cost the insurance company between $5,000 and $15,000.

The distribution of the health disturbance is estimated non-parametrically using the residuals of the health production function. We assume that $\mathbb{E}[\epsilon^h_{it}|X^h_{it}] = 0$, where X^h_{it} is the vector of regressors in the health production function.

Survival: At the end of any period t individuals may survive into the next, denoted by $S_{it+1} = 1$, with the following probability

$$D_{it+1}\left(z_{it+1}\right) \equiv \Pr[S_{it+1} = 1|z_{it+1}] = \frac{1}{1 + \exp(X^d_{it}\Gamma^d)} \tag{17}$$

where $X^d_{it} = [1, h_{jit}, \ldots, h^5_{jit}, a_{it}, a^2_{it}, b_i, edu_i, x_{jit}]$.

3.1.5 The Value Function

We define the value function conditional on choice $j \in \mathcal{C}_{it}$, net of taste shocks, for individual i at time t as follows:

$$v_{jit}(z_{it}) = \mathbb{E}\left[y_{jit} + \beta \left[D_{it+1}\left(z_{it+1}\right) \max_{c \in \mathcal{C}_{it+1}} \left\{ v_{cit+1}(z_{it+1}) + \varepsilon_{cit+1} \right\} \right] \middle| z_{it}, j \right] \tag{18}$$

Expectations are taken over product characteristics affecting the flow utility and the evolution of both observed and unobserved state variables. Expectations over the evolution of unobserved state variables are independent conditional on the current set of state variables. Therefore, we can rewrite equation (18) as

$$v_{jit}(z_{it}) = \mathbb{E}_y[y_{jit}|z_{it}] + \beta \mathbb{E}_z \left[D_{it+1}\left(z_{it+1}\right) \mathbb{E}_\epsilon \left[\max_{c \in \mathcal{C}_{it+1}} \left\{ v_{cit+1}(z_{it+1}) + \varepsilon_{cit+1} \right\} \right] \middle| z_{it}, j \right] \tag{19}$$

The first expectations operator, $\mathbb{E}_y$, denotes expectations over outcomes that affect flow utility, including income and physical ailments. The second operator, $\mathbb{E}_z$, denotes expectations over the evolution of observed state variables z_{it}, including health and variables that summarize the state of the product market. The third operator, $\mathbb{E}_\epsilon$, denotes expectations taken over the joint distribution of unobserved choice-specific taste shifters.

3.2 The Evolution of the Aggregate State

Until now, we have described the individual's decision conditional on the available set of products and other aggregate characteristics of the market. Now we turn to the evolution of aggregate market-level characteristics, which includes product entry and exit. The set of market-level state variables, denoted $z^{\mathcal{M}}$, contains the characteristics of all market products that are available at time t, W_t, a summary of the current state of technology, ω_t, and the

aggregate joint distribution of individual characteristics, $\mathcal{F}_t$. In the model, the number of new products is a function of previous innovations and the share of trial users. The characteristics of new products are a function of aggregate behavior summarized by ω_t and an exogenous distribution of innovations. Product exit is also a function of individuals behavior: products are leave the market if their shares determined by aggregate consumer choices fall below a threshold, where the threshold is explained below. Finally, since aggregate consumer choices (a function of state variables) help to determine the evolution of the market, each individual who forecasts product evolution must also form expectations over aggregate state variables of the consumers in the market.

3.2.1 Product Entry

In each period, entry of new products occurs according to a reference point for innovation or *centroid*, denoted ω_{t-1}, a distribution of characteristics of new products $F_{\theta|\omega_{t-1}}$ and a distribution of number of new products F_N.

Centroid: (ω_t): At any period t, the centroid for innovation is a weighted average of products available last period, given by:

$$\omega_t = \sum_{k \in \mathcal{P}_{t-1}} \left(s^P_{t-1} \left(s_{k|P,t-1} \right) + \sum_{g=1}^{\mathcal{G}_{t-1}} 1\left\{ k \in g \right\} s^g_{t-1} \tilde{s}_{k|g} \right) \theta_k \tag{20}$$

where θ_k are the characteristics of product k. The weight given to the characteristics of product k is the sum of the lagged share of people "staying" on their previous product, s^P_{t-1}, multiplied by the share of stayers who were taking product k, $s_{k|P,t-1}$, plus the sum over all groups of an indicator of product k belonging to group g multiplied by the lagged share of people choosing to switch into cluster g, s^g_{t-1}, times the weight of product k in cluster g, $\tilde{s}_{k|g}$. The lagged shares s^P_{t-1} and s^g_{t-1} are conditional on consumption of a market product so that $s^P_{t-1} + \sum_{g=1}^{\mathcal{G}_{t-1}} s^g_{t-1} = 1$.

Characteristics of New Products $(F_{\theta|\omega_t})$: Every new product introduced at t, characterized by θ, is an innovation around the previous-period centroid:

$$\theta = \omega_{t-1} + \nu \tag{21}$$

where $\nu \sim F_v$ is the stationary distribution of innovations. Notice that ω_{t-1} and F_v determine $F_{\theta|\omega_{t-1}}$.

Number of New Products (F_N): In each period, we observe that the number of new products introduced in the market varies. Moreover, the number of products thrown into the market seems to be related to the size of previous discoveries as well as to the share of individuals consuming the trial product. We capture these facts in our specification for the distribution over the number of draws to be taken from $F_{\theta|\omega_{t-1}}$. At any period, a number New_t of new products may enter the market at t. This number follows a negative binomial that permits dispersion in the mean:

$$\begin{aligned} New_t|_{\mu^*} &\sim Poisson\left(\mu_t^*\right) \\ \mu_t^* &\sim Gamma\left(1/\alpha^N, \alpha^N \mu_t\right) \\ \mu_t &= \exp(\beta_0^N + \beta_1^N MaxChange_{t-1} + \beta_2^N TrialsShare_{t-1}) \\ \ln \alpha^N &= \alpha_0^N + \alpha_1^N MaxChange_{t-1} \end{aligned} \tag{22}$$

where α^N and β^N are vectors of parameters to be estimated. The binomial model is conditioned on two covariates: $MaxChange_{t-1}$ and $TrialsShare_{t-1}$. The share of people going to trials in the previous period captures the fact that more experiments can be advanced if more people participate in clinical trials. The variable $MaxChange_{t-1}$ captures the relatively higher number of new products that follow the appearance of better innovations. It is computed as follows:

$$MaxChange_{t-1} = \sum_{r \in h,x} \frac{\max_{\theta^r \text{ new at } t-1} \left\{\theta^r - \omega_{t-2}^r\right\}}{\max_{\theta^r \text{new at } \tau, \forall \tau} \left\{\theta^r - \omega_{\tau-1}^r\right\}} \tag{23}$$

It measures the distance between the previous period's new products and the previous period's centroid. The relative change is computed for each of the two characteristics (health and no-ailments) and is scaled by the maximum change observed over the sample period.[15] The specifications of ω_t, $F_{\theta|\omega_{t-1}}$, and F_N render the path of innovation endogenous. The reason is that individual choices, summarized by market shares, affect the centroid in equation (20). By affecting ω_t, market shares affect the characteristics of every new product θ in equation (21). Intuitively, treatments that keep patients alive and those associated with fewer ailments will capture larger shares of the market and firms will innovate on drugs with larger market shares. Additionally, individuals' choices affect the path of innovation through

[15]Note that in order to compute $MaxChange_{t-1}$ we need the scaling quantities given by

$$\max_{\theta^r \text{ new at } \tau, \forall \tau} \left\{\theta^r - \omega_{\tau-1}^r\right\} \tag{24}$$

for $r \in \{h, x\}$ which are estimated consistently by their data counterparts. Recall all new products at t are draws from the innovations distribution conditional on ω_{t-1}.

their effect on the distribution of number of new products.

3.2.2 Product Exit

Incumbent drugs may also exit the market. Exit happens on the underlying product space $\mathcal{P}_t$ and it happens at two different levels: exit for switchers and overall exit. Exit for switchers happens when the product is no longer available for people to switch into it but people may still stick to it if they consumed the product in the prior period. Individuals face regulation that prevents them from buying the product they consumed last period and selling it to other customers that wish to switch to it. Overall exit happens when the product is no longer available to any consumer. Exit happens according to the following rules that aim to reconcile empirical observation and theory—where expected shares must be positive due to assumptions on the taste shocks.

1. If the ratio of people switching and being assigned product k relative to the number of people switching falls bellow $\tilde{\sigma}_1$ during three consecutive periods, the product is withdrawn from the market. $\tilde{\sigma}_1$ is chosen as the minimum conditional share observed in the data and the number 3 is chosen to smooth the market spells of products.

2. If the ratio of people consuming product k, either by staying or switching, relative to the number of people consuming a market product falls below $\tilde{\sigma}_2$ during two consecutive periods, the product is withdrawn from the market. $\tilde{\sigma}_2$ is chosen as the minimum conditional share observed in the data and the number 2 is chosen to smooth the market spells of products.

3.2.3 Aggregate Consumer Characteristics

In order to forecast aggregate behavior that determines the path of technology, individuals form expectations over the joint distribution of individual characteristics, $\mathcal{F}_t$. In section 3.1.1 we specified individuals characteristics to be race, education, health, age, working status, person-specific income characteristics, and characteristics of last product consumed. Conditional on the current aggregate state, the future aggregate state is simply a mapping of aggregate behavior. Formally,

$$\mathcal{F}_{t+1} = \mathcal{G}\left(\mathcal{F}_t, G(\varepsilon), \omega_t, W_t\right) \tag{25}$$

$\mathcal{G}$ is the mapping from today's aggregate state into one-period-ahead aggregate state implied by current period choices and states.

Consider the following example to fix ideas. Suppose the only individual characteristic in the support of $\mathcal{F}$ is health. Then, equation (16) reduces to

$$h_{it+1} = \widetilde{f}(h_{it}, d_{it}(h_{it}, \varepsilon_{it}, \omega_t), \omega_t) + \epsilon_{it}^h$$

for some function $\widetilde{f}$, where d_{it} is the vector representing the individual's choice. Let f_{ϵ_h} be the pdf of health shocks. Then the probability that an agent chosen at random will have health h' tomorrow can be forecasted as

$$\begin{aligned}\Pr(h_{t+1} = h' | \mathcal{F}_t(h), G(\varepsilon), \omega_t) &= \int f_{\epsilon_h}\left(h' - \widetilde{f}(h, d(h, \varepsilon, \omega_t), \omega_t)\right) d\mathcal{F}_t(h) dG(\varepsilon) \\ &= \int f_{\epsilon_h}\left(h' - \widetilde{\widetilde{f}}(h, \varepsilon, \omega_t)\right) d\mathcal{F}_t(h) dG(\varepsilon)\end{aligned}$$

4 Estimation

We start by summarizing our estimation procedure. We then provide more details about some of the pieces involved. For a more extensive treatment of the estimation procedure we direct the reader the estimation appendix (Appendix B). Our estimation procedure can be summarized in the following steps

1. *Definition of products as treatments.* Our estimation starts with the definition of products. We define a product as a combination of single-product components. Examples of products are AZT or the combination of AZT+3TC+Saquinavir.

2. *Estimation of outcome equations.* We estimate processes for income, out-of-pocket payment, labor supply and survival. Heath and no-ailments equations will be estimated in the next step (see equations (11), (12), (13), and (17)).

3. *Estimation of product characteristics.* Given products defined in step 1, we estimate product characteristics (see equations (34) and (35)).

4. *Clusters.* Using the estimated product characteristics in step 3, we use a k-means algorithm to obtain clusters of products for every period (see equations (4)).

5. *Within cluster weights.* Using the clusters obtained in step 4 and the product characteristics from 3, we obtain within clusters weights as non-linear regressions of within cluster share on covariates (see equations (6) and (7)).

6. *Cluster characteristics.* Using the clusters obtained in step 4, product characteristics from step 3, and within cluster weights form step 5, we compute cluster characteristics—mean and variance matrix.

7. *Centroid.* Using product characteristics from step 3, clusters from step 4, and within cluster weights from step 5, we back out innovation centroids for every period (see equation (20)).

8. *Distribution of innovations.* Every new product is modeled as a draw around the centroid (see equation (21)). Hence, for every new product at a given period we compute the realized innovation around the centroid, which is the residual from subtracting the centroid (step 7) from the product characteristic (step 3). Using the realized innovations we non-parametrically estimate the stationary distribution of innovations, F_v.

9. *Distribution of number of draws.* Using data regarding the amount of new products per period we estimate the distribution of number of new products specified as a negative binomial with dispersion in the mean (see equations (22) and (23)).

10. *Conditional choice probabilities.* Using cluster characteristic from step 6, centroids from step 7 and other aggregate and individual-specific state variables we estimate parametric conditional choice probabilities (see equations (36), (38), and (37)).

11. *Structural utility parameters.* We follow Hotz et al. (1994) and use forward simulation to generate choice and technology paths as well as future individual states that will serve as inputs to the simulated future value function. In our forward simulation we use estimated conditional choice probabilities (step 10), the distribution of number of draws (step 9) and the distribution of innovations (step 8) as well other estimated processes (step 2 through step 7). Finally, we implement a GMM estimator using a moment condition which is a function of the forward simulated data, conditional choice probabilities, and utility parameters.

In section 4.1, we derive theoretical moment conditions and their empirical counterparts used in estimation. In Section 4.2, we discuss our forward simulation.

4.1 The Moment Conditions

We obtain two different representations of the differences in conditional value functions to construct moment conditions that we use to estimate the structural utility parameters. Our

first representation follows from our assumption that the taste shocks ε_{jit} are iid Extreme Value Type-I. Under this assumption, for any $j, o \in \mathcal{C}_{it}$ we know that the difference in conditional value functions can be written as

$$v_{jit}(z_{it}) - v_{oit}(z_{it}) = \ln\left(\frac{p_{jit}(z_{it})}{p_{oit}(z_{it})}\right) \tag{26}$$

Alternatively, we can write the conditional value function as the expected value of the sum of current and future flow payoffs conditional on optimal behavior. This representation is presented in Proposition 1

Proposition 1. *Let $V(z_{it}, \varepsilon_{it})$ be the value function for individual i at period t who has a state given by z_{it} and ε_{it}. Define $P_j^{o(s-1)}$ as the probability of surviving until period $t+s-1$ conditional on the state at t, decision j at t, and optimal behavior, denoted d_i^o, up to some period $T^* > t$.*[16] *Define $\psi_{kit}(z_{it}) \equiv \mathbb{E}_\varepsilon[\varepsilon_{kit} | d_{it}^o = k, z_{it}]$ as the expected value of the kth taste shock conditional on alternative k being optimal. Finally, let γ be the Euler constant. Then, the conditional value function can be written as*

$$\begin{aligned} v_{jit}(z_{it}) &= \mathbb{E}[y_{jit} | z_{it}] + \sum_{s=1}^{T^*} \beta^s P_j^{o(s-1)}(z_{it}) \times \\ &\mathbb{E}_z\left[D_{it+s}(z_{it+s}) \sum_{k \in C_{t+s}} p_{kit+s}(z_{it+s}) \left[y_{kit+s}(z_{it+s}) + \psi_{kit+s}(z_{it+s}) \right] \middle| z_{it}, d_{jit} = 1, S_{it+s-1} = 1, d_i^o \right] \\ &+ \beta^{T^*+1} P_j^{o(T^*)}(z_{it}) \, \mathbb{E}_z\left[D_{it+T^*+1}(z_{it+T^*+1}) \, V(z_{it+T^*+1}, \varepsilon_{it+T^*+1}) | \, z_{it}, j, S_{it+T^*} = 1, d_i^o \right] \end{aligned} \tag{27}$$

and

$$\psi_{kit}(z_{it}) = \gamma - \ln(p_{kit}(z_{it})) \tag{28}$$

Proof: see Appendix C

We then choose a value of T^* high enough so that the product $\beta^{T^*+1} P_j^{o(T^*)}(z_{it})$ approaches zero, eliminating remaining differences in conditional value functions. This yields our second representation for the difference in conditional value functions

$$v_{jit}(z_{it}) - v_{oit}(z_{it}) \approx \mathbb{E}[y_{jit} - y_{oit} | z_{it}] + \bar{v}_{jit}(z_{it}) - \bar{v}_{oit}(z_{it}) \tag{29}$$

[16] Since any individual present at t has evidently survived until t, $P_j^{o(0)}(\cdot) \equiv 1$.

where

$$\begin{aligned}\bar{v}_{jit} &= \sum_{s=1}^{T^*} \beta^s P_j^{o(s-1)}(z_{it}) \times \\ &\mathbb{E}_z\left[D_{it+s}(z_{it+s}) \sum_{k \in C_{t+s}} p_{kit+s}(z_{it+s}) \left[y_{kit+s}(z_{it+s}) + \psi_{kit+s}(z_{it+s}) \right] \middle| z_{it}, d_{jit} = 1, S_{it}^{(s-1)} = 1, d_i^o \right]\end{aligned} \tag{30}$$

Let $J = 6$ be the maximum possible cardinality of the individual's choice set. For every individual at every period we have at most $J-1$ differences of the form displayed in equation (29). Let $w(z_{it})$ be a vector of instruments orthogonal to the difference between the alternative representations of the difference in conditional value functions. Therefore, we can form the following moment conditions

$$\mathbb{E}\left\{ w(z_{it}) \otimes \begin{bmatrix} \ln\left(\frac{p_{oit}(z_{it})}{p_{1it}(z_{it})}\right) + \mathbb{E}[y_{1it} - y_{oit}|z_{it}] + \bar{v}_{1it}(z_{it}) - \bar{v}_{oit}(z_{it}) \\ \vdots \\ \ln\left(\frac{p_{oit}(z_{it})}{p_{J-1it}(z_{it})}\right) + \mathbb{E}[y_{J-1it} - y_{oit}|z_{it}] + \bar{v}_{J-1it}(z_{it}) - \bar{v}_{oit}(z_{it}) \end{bmatrix} \right\} = 0 \tag{31}$$

As econometricians, we observe the underlying stochastic process that gives rise to the stochastic process of cluster characteristics that agents observe. Therefore, we compute the expectation in equation (31) with respect to the underlying stochastic process denoted by $\mathcal{P}_t$. This is crucial for our simulation estimation method. The key fact is that we observe the characteristics of the underlying process of product evolution and that allows us to simulate the stochastic evolution of clusters. This procedure is valid as

$$\mathbb{E}\left\{ w(z_{it}) \otimes \mathbb{E}_z[\cdot] | \mathcal{P}_t \right\} = \mathbb{E}\left\{ w(z_{it}) \otimes \mathbb{E}_z[\cdot | \mathcal{P}_t] \right\} \tag{32}$$

We simulate choices, transitions, and technology paths to obtain a version of the right hand side of equation (29) where expectations are taken conditional on the underlying stochastic process $\mathcal{P}_t$ that generates the cluster characteristics agents observe. We then form sample analogs of the moment conditions on (31) and estimate parameters using generalized method of moments.

4.2 Forward Simulation and Conditional Choice Probabilities

We use forward simulation to obtain an expression for the difference in conditional value functions. A key component of our forward simulation procedure is that innovation is endogenous to individuals choices. To understand how this works, begin with an individual i

at time t with state variables $\mathcal{Z}_{it}$ facing choice set $\mathcal{C}_{it}$. Using parametric conditional choice probabilities estimated beforehand we simulate his choice at time t. Next, using estimated parameters for the stochastic processes governing outcomes and transition, we simulate his state variables at $t+1$. We then simulate $t+1$ choices conditional on the new simulated state and continue the same process until we reach T^*.

Innovation is endogenous to consumer choices. Therefore, in order to simulate expectations for individual i at period t we must simulate an entire artificial technological path—that is specific to him—and respective choices for all individuals available in period t. Implied market shares in this simulated world determine the future simulated centroid for innovation.

Although we as econometricians know what products the underlying stochastic process generates, individuals only observe the stochastic process of cluster characteristics. We simulate the underlying stochastic evolution of products and obtain from there the state variables that are relevant for consumer choices.

4.2.1 Product Characteristics and Number of New Products

We estimate product characteristics using data on individual treatment usage and subsequent reports of health and ailments. Our estimation equations mimic equations (15) and (16), which individuals use to form expectations over their health and ailments conditional on their choice. The key difference between equations (15) and (16) and our estimation equations is that here our aim is to obtain characteristics of each individual treatment.

Let δ_{rit} be an indicator that treatment r was used by individual i at time t. The characteristics of treatment r are denoted

$$\theta_r = \left\{\theta_r^x, \theta_r^h\right\} \in \mathbb{R}^2 \tag{33}$$

The components of θ_r are estimated as the coefficients of δ_{rit} in the health and no-ailments regressions

$$h_{it} = \sum_{m=0}^{5} \alpha_m^h h_{it-1}^m + \sum_r \theta_r^h \delta_{rit} + \epsilon_{it} \tag{34}$$

$$\Pr\left[x_{it} = 1|\cdot\right] = \frac{\exp\left(\sum_{m=0}^{5} \alpha_m^x h_{it-1}^m + \sum_r \theta_r^x \delta_{rit}\right)}{1 + \exp\left(\cdot\right)} \tag{35}$$

Finally, the distribution of number of new products is estimated using the specification in (22) and data on product entry. We now turn to the empirical distribution of innovations.

4.2.2 The Distribution of Innovations

The characteristics of new products entering the market today are determined by last period's centroid and a draw from the distribution of innovations, F_ν. This distribution, which we assume is stationary, provides the distance of new products relative to the current centroid. In order to construct the distribution of innovations we use all periods available in the MACS data with relevant information on treatment consumed, health, and ailments. These data span from 1986 to 2008. We then define centroids for innovation, ω_t, given by equation (20). For each new product at t, characterized by θ, we compute a realized innovation vector as

$$\nu^\theta = \theta - \omega_{t-1}$$

There are 76 realizations from the innovation distribution which we use to obtain a nonparametric empirical distribution for ν.

4.2.3 Conditional Choice Probabilities

The probability that an individual chooses one of the alternatives depends on the elements of his state. As such, the conditional choice probabilities needed to simulate choices in our estimation method are functions of individual-specific variables as well as market-level variables.

Individuals decide between one of $\mathcal{G}$ clusters, yesterday's product (if any), a trial product, and no product. Let W_{jit} be the characteristics describing alternative j for individual i at period t: mean health, mean ailments, and the variance matrix. Let $W_{jit}W_{jit}$ denote a vector of interactions between the elements of W_{jit}. Let $\tilde{x}_{it}$ and $\tilde{z}_{it}$ be subsets of the individual-specific components of the state.[17] Let $\omega_t W_{jit}$ denote a vector of interactions between the centroid and the elements of W_{jit}. Similarly, let $W_{jit}\tilde{z}_{it}$ be a vector of interactions between the components of W_{jit} and individual-specific state components and let $\omega_t W_{jit}\tilde{z}_{it}$ be defined in a similar fashion. Finally, let $\tilde{\mathcal{F}}_t$ denote a set of non parametric moments describing the joint distribution of aggregate characteristics, $\mathcal{F}_t$.[18]

For each of the alternatives, the conditional choice probabilities (ccps) are expressed as follows:

[17] $\tilde{z}_{it}$ includes h_{it-1}, a_{it-1}, b_i, l_{it} while $\tilde{x}_{it}$ includes a constant, a_{it-1}, b_i.

[18] We specify these moments as shares of people with different sets of characteristics.

Cluster ccps $(j = 1, \ldots, \mathcal{G})$

$$p_{jit} = \frac{\exp\left(\gamma_0 \tilde{x}_{it} + \beta_0 W_{jit} + \beta_1 W_{jit} W_{jit} + \beta_2 \omega_t W_{jit} + \beta_3 W_{jit} \tilde{z}_{it} + \beta_4 \omega_t W_{jit} \tilde{z}_{it} + \beta_5 W_{jit} \tilde{\mathcal{F}}_t\right)}{1 + \sum_{k=1}^{\mathcal{G}+2} \exp(\cdot)} \tag{36}$$

γ_0 is constant across clusters and over time. For a given cluster j and period t, W_{jit} is in fact constant across individuals so $W_{jit} = W_{jt}$.

Trial ccps $(j = \mathcal{G} + 1)$

$$p_{jit} = \frac{\exp\left(\gamma_j \tilde{x}_{it} + \beta_0 W_{jit} + \beta_1 W_{jit} W_{jit} + \beta_3 W_{jit} \tilde{z}_{it} + \beta_5 W_{jit} \tilde{\mathcal{F}}_t\right)}{1 + \sum_{k=1}^{\mathcal{G}+2} \exp(\cdot)} \tag{37}$$

For the trial alternative, W_{jit} is constant across individuals so $W_{\mathcal{G}+1it} = W_{\mathcal{G}+1t}$. In fact, two of the components of W_{jt} are $\omega_{t-1} + \mu_\nu$, where μ_ν is the mean of the innovations distribution. Therefore, to avoid collinearity we do not include terms $\omega_t W_{jt}$ and $\omega_t W_{jt} \tilde{z}_{it}$ in the trials ccps.

Staying ccps $(j = \mathcal{G} + 2)$

$$p_{jit} = \frac{\exp\left(\gamma_j \tilde{x}_{it} + \beta_0 W_{jit} + \beta_1 W_{jit} W_{jit} + \beta_2 \omega_t W_{jit} + \beta_3 W_{jit} \tilde{z}_{it} + \beta_4 \omega_t W_{jit} \tilde{z}_{it} + \beta_5 W_{jit} \tilde{\mathcal{F}}_t\right)}{1 + \sum_{k=1}^{\mathcal{G}+2} \exp(\cdot)} \tag{38}$$

When individuals choose to stick to their previous product $W_{\mathcal{G}+1it}$ becomes heterogeneous—individuals may have consumed different products last period.

No product ccps $(j = 0)$

$$p_{jit} = 1 - \sum_{k=1}^{\mathcal{G}+2} p_{kit} \tag{39}$$

Even though the characteristics of the choice set are non stationary, by interacting our time-varying regressors $\tilde{z}_{it}$ with the characteristics of the choice for individual i, W_{jit}, we are able to control for the state of the world inside the ccps. As a consequence of this we do not have to run period-specific logits and we can have ccps for any simulated world as long as our observed worlds cover the space of possible worlds reasonably well. We also include parameters that are invariant to the state of the technology, γ, which capture stationary taste differences between staying in current choice, trying a new market product, going to a trial, or not consuming anything. Also, since all clusters correspond to the action of "trying a market product" we impose $\gamma_j = \gamma_{j'} = \gamma_0$ for any $j, j' = 1, \ldots, \mathcal{G}$.

5 Parameter Estimates and Choice Dynamics

In this section, we discuss estimates of the structural model. We organize our discussion around the key factors driving choices. First, individuals form beliefs about the distribution of innovations (Section 5.1). Second, they obtain flow utility (Section 5.2). Third, they consider how each treatment implies a different distribution over outcomes and future states (Section 5.3). Taken together, we can then use the model to simulate choices over time, which allows us to assess model fit. (Section 5.4).

5.1 The Distributions of Innovations and New Products

In our model, every new product is an innovation about the centroid. How far new products land from the centroid is stochastically given by the distribution of innovations, F_ν. Using equation (21) and estimated product characteristics we can back out the innovation implied by each product as the difference between its characteristics and the centroid.

Figure 5 shows our non parametric estimate of F_ν. The distribution of innovations is bimodal and it does not appear to be well approximated by a standard parametric distribution. One of the modes is located approximately at the status quo point $(0, 0)$. A second mode is located north of the first mode along the health axis into the region where changes on health quality are positive. This yields a distribution that has a positive mean in terms of health quality and a mean that is approximately zero for no-ailments quality.

The estimated distribution of innovations suggests that, if draws were random, on average products would improve over time in terms of health quality but would remain largely unchanged in terms of no-ailments quality. In our model, however, the moments of the innovations distribution interact with consumer demand to shape the path of technology. Future products are drawn more often from parts of the distribution where market shares are larger, which could shift the path of innovation. In this sense, innovation is endogenous to market shares, themselves a function of optimal consumer choices.

Estimates for the distribution of number of new products are shown in Table 3. Results show that the magnitude of previous innovations increases the likelihood that more new products enter market. Further, estimates mean that large positive innovations are likely to be followed by the appearance of a multitude of products, which is consistent with firms vying for market share following a breakthrough. The magnitude of previous innovations also reduce sthe dispersion around the number of new products that enter. The share of consumers opting for the trial product in the prior period also increases the likelihood of more products entering the market. The reason is that, as more consumers select trial products.

firms have more room for experimentations which provides them with valuable information about the viability of new treatments that they can now introduce into the market more rapidly. The fit of our distribution of new products is shown in Figure 6. It shows that the empirical distribution is not far from the average over time of the predicted probabilities generated by the model.

5.2 Utility Parameter Estimates

Utility parameters are reported in Table 4. Recall, we interact patient socio-demographic and health characteristics with the utility they gain from each treatment choice, where choices include drawing from a cluster, entering a trial or staying on the current treatment (where there is no uncertainty about product quality. These interactions help to explain heterogeneity in choices across groups that are not attributable to variation in continuation payoffs. In interpreting parameter estimates, note that the utility from no treatment is normalized to zero across groups. Therefore, parameter estimates govern period utility for different groups relative to what they gain from taking no treatment.

According to parameter estimates clusters and trials lead to a utility cost and, generally, these penalties are higher for non-white patients. Black men face a particularly high penalty of trial participation, a finding that is consistent with a broad literature investigating historical reasons why blacks are reluctant to enter trials to use experimental drugs. Moreover, healthier individuals have a lower utility of treatments where they face uncertainty, including clusters or in trials. Interestingly, healthier individuals gain utility from using drugs they have used before. These results suggest that healthier individuals dislike uncertainty about drugs and, perhaps, switching costs relative to their less healthy counterparts. We also find that the utility costs of treatment relative to no treatment are stronger for younger individuals. This is perhaps reflective of age-dependent tolerance for medication, especially if older individuals have grown accustomed to using medications for other health problems.

Finally, individuals dislike ailments regardless of which product they are using. This utility parameter is key as it explains why individuals eschew medications that have high dynamic payoffs in the form of better future health. This finding is consistent with Chan and Hamilton (2006) and Papageorge (2016) in the context of HIV, which has demonstrated that even in the context of a deadly infection, individual treatment choices reflect a distaste for side effects. Finally, the utility function shows that individuals gain positive utility from income, which reflects consumption utility and is expected.

5.3 Transitions and Outcomes

Next, we discuss the processes describing how state variables produce outcomes or transitions to other states. The quality of any product r, θ_r, has two dimensions: health, which we estimate using equation (16), and no-ailments, which we estimate using equation (15). Estimated product characteristics are product-specific coefficients in these regressions. To conserve on space, we present coefficient estimates in Tables S5 and S6, found in Appendix D (see Column 5 in both tables). These are the coefficients that patients take into account when deciding among treatment options. However, We can also see how product quality evolves over time in Figure 4.

Individuals in our model care not only about health and ailments but also about a number of other outcomes which are affected by them: income, out-of-pocket payments, labor, and survival. Tables 5 presents our results for the income equation. Income increases with health. Moreover, individuals who do not suffer ailments also have higher income as their productivity is likely to be higher. Minorities have lower income. Income is concave in age and it increases with labor participation and education. At any period individuals may incur in out-of-pocket payments not uniquely related to their treatment consumption decision. According to Table 6, conditional on having out-of-pocket expenditures, these payments increase with age. Minorities spend less and more educated people spend more. Similarly, individuals that suffer ailments spend more, perhaps because they are managing other health conditions. Even with heavy subsidization in the HIV treatments market, individuals wanting to consume must still pay part of the cost and this is reflected in higher expected payments. Labor market participation increases expected payments, which may reflect different pricing schemes for public versus private insurance.

Labor participation is stochastic in our model and it is revealed to individuals at the beginning of the period. Estimates in Table 7 shows that the log odds ratio of working versus not working increases with age until about age 40 and then decreases. Odds of working increase with education and it increases substantially if the individual had worked the previous period. At the end of every period individuals face the possibility of death. Estimates in Table 8 imply that the log odds ratio of death versus survival decreases with age until about age 35 and then increases. The likelihood of death is smaller for black individuals and for individuals who are not suffering ailments.

Health plays a major role in our model and also exhibits strongly non-linear relationships with other outcomes, which help to explain differences in optimal choice for individuals with somewhat similar health profiles (as measured by CD4 count). Therefore, in presenting results on outcomes, we show how they relate to health according to model estimates. In

particular, in Figure 7, we plot the relationship between health and several outcomes: income, out-of-pocket payments, labor supply and survival. According to the figure, income increases steeply with CD4 count for very sick individuals but the effect of health flattens substantially for individuals with CD4 counts above 250. The health profile of out-of-pocket payments in Figure 7 is the mirror image of the health profile for income with deeper decreases in payments as health increases for the sickest. This makes sense as health expenditures due to opportunistic infections, for example, would be expected to decline precipitously as a result of small health increases at low health levels. Similarly, the odds of working increases with health until a CD4 count of about 350 units and then it flattens. Finally, the effect of health increases on survival are more dramatic the more sick individuals are. Even though the positive impact of health on survival remains at higher health levels, this relationship diminishes considerably after a CD4 count of about 250 units.

In general, the health profiles in Figure 7 tell a very consistent story about CD4 count and HIV infection. The effect of marginal health increases on outcomes is much stronger for individuals with low CD4 counts and it seems to flatten after individuals surpass well-known cutoffs below which AIDS occurs. This is consistent with the idea that lower CD4 counts have little discernible impact on symptoms or survival until the AIDS threshold is reached. Below that threshold, further reductions have large effects on outcomes since the body's immune system becomes increasingly compromised and is therefore unable to fight off routine infections. These results highlight how it is not appropriate to model health as having a linear effect on outcomes in the context of HIV. Rather, CD4 count matters insofar as it affects outcomes, which is most likely to happen at very low levels.

5.4 Simulated Choice Dynamics and Model Fit

In Figure 8, we plot observed treatment choices over time along with those generated by the model. Note that ccp estimation fit is discussed in Appendix B. In general, we are able to capture basic trends, including the rise in treatment usage as drugs improve through innovation. We also capture trials participation dynamics fairly well. The model does not do a very good job of reproducing the spike in participation shortly before HAART introduction. The reason is that the model only accounts for changes in the demand for trials, which occurred if individuals anticipated an innovation for which they sought early access. However, demand shifts are only one part of the story; there was also a shift in the supply of trials as a number of new drugs were tested that would eventually comprise HAART. If so, the spike in participation and would not be fully captured by the our model

that focuses on patient demand.[19] Beyond this spike, however, our model can capture the main contours of choice dynamics.

6 Counterfactual Policy Simulations

In this section, we discuss counterfactual policy simulations. The policies we evaluate center around demand externalities. In our framework, the observed path of innovations can be viewed as a draw from a distribution of innovations, itself a function of optimal consumer choices. This feature of our model means that we can explicitly relate optimal choices to the distribution of innovations. In this section, we begin by illustrating how the model we estimate generates a distribution of innovations at each point in time. This means we can assess the probability of the observed path of innovations and ask, for example, whether the observed path included tail events (Section 6.1). Having established that our model generates a distribution of possible innovations, we go on to discuss how different kinds of choice dynamics would influence the distribution. We demonstrate that alternative market shares could speed or slow the development of technologies that potentially increase social welfare (Section 6.2). This naturally leads to a discussion of policies that could raise consumer welfare by subsidizing some choices over others. We focus on the idea policies affecting consumption choices could potentially solve a coordination problem whereby consumers under-experiment and therefore slow the progress of technology (Section 6.3). As we show, though it may be individually rational to underuse such technologies, the question arises whether it is socially optimal to induce all consumers to use technologies that spur innovation.

6.1 The Distribution of Technology Paths

Imbedded in our estimation procedure is the simulation of different innovation paths. Alternative paths are drawn from the same innovation distribution from which the realized path is drawn. This means that we can assess the realized path of distributions in comparison to the full distribution of paths. In particular, we take the 1990 distribution of state variables as given and then simulate technology, choices and state variables for 18 years. We do so again for the 1997 distribution of state variables. For each simulated period, previous period market shares and the current distribution of technology are used to simulate draws of new

[19] In a companion paper, we model supply of trials more explicitly and demonstrate the the increase in trials prior to the introduction of HAART in part explains the observed spike in the likelihood of participation. In the current framework, we could model supply shifts in a reduced-form manner as a temporary decrease in the utility cost of joining a trial, which would reflect the ease of finding a trial in which to participate. We abstract from supply here, however, since the focus of our model is on demand shifts and innovation.

products. Consumers optimally choose from among existing new products, which generates one-period-ahead market shares and also a probability distribution over one-period-ahead states, from which we draw transitions. This leaves is with a one-period ahead set of consumers and products and the process is repeated. We plot the simulated distribution of outcomes across time and then compare it do the data.

First, we consider aggregate health and the health centroid if we begin the simulation in either 1990, prior to HAART introduction, and in 1997, once HAART has been introduced. Results are plotted in Figure 9, where the green line is the realized path, the black line is the mean simulated path and the dotted lines are confidence intervals. Considering the plots on the left, where the simulation begins in 1990, it is clear that HAART introduction was a tail event. The observed path of innovations follows the simulated paths quite well until 1996. Thereafter, the health centroid, which summarizes effectiveness of market drugs, along with aggregate health of consumers in the market is far above what would have been expected. Between the years 1996 and 2000, the realized path of the centroid is outside of the 95-percent confidence interval. Interestingly, the expected centroid approaches the realized centroid as time goes on. This means that the gradual progress of technology would have eventually been expected to improve drug effectiveness until something nearly as good as HAART would have come along, though far later than it actually did. Looking at the right side of Figure 9, where the simulation begins in 1997, notice that the realized path underperforms the average simulated path. This means that technology, measured by effectiveness, was expected to improve more than it did from the perspective of 1997.

In Figure, 10, we perform a similar analysis for the other product quality: lack of side effects. Here, realized product quality (measured by the centroid and by aggregate outcomes) seems to have underperformed what would have been expected from the distribution of innovations. In fact, one of the disappointments with regard to HAART is that its side effects were quite harsh, which led many HIV+ men to avoid using it despite its effectiveness (Papageorge, 2016). In Figures 11, we consider survival and consumption. The results on both match those on health: HAART introduction was a tail event, which increased survival and product consumption. Finally, in Figure 12, we compare simulated paths with the realized path, considering product entry and exit. We under-estimate number of product. When HAART is introduced, there is a huge number of new entries (tail event). Afterwards, we find that new products beget new products, so that even though the rate is in line with simulations, the total number is too low. Note, our under-estimation of entry does not mean that our model is a poor fit. In fact, it this shows that our model is successful at treating breakthroughs (and subsequent entry of products and behavior of firms) as tail events.

6.2 Demand Pull: How Consumer Choices Affect Innovation

Findings from this section demonstrate that a policy that changes consumer choices will affect the path of innovation. Notice, this means that there are demand externalities and that a social planner could improve welfare, which we discuss next. Here, we consider technology and aggregate outcomes under different choice policies. In the first, individuals are dynamic optimizers. In the second, the are myopic. In the third, choice is random.

In Figure 13, we consider average health and ailments under each regime (left-hand-side plots) and centroid health and ailments. In Figure 14, we consider average survival, number of consumers, entry and exit under the different choice regimes. Consider average health. It rises if agents choose products randomly. This means that individuals under-use products that would improve technology and lead to higher average health in the future. This can by seen with the health centroid in the right-hand-side panel. The likelihood of no ailments, however, is lower in the static model. This may seem puzzling since consumers in the static model only care about side effects. This finding reflects how improved health affects ailments in the following period, not through side effects, but from fewer symptoms. Therefore, if agents are myopic and only care about side effects in the current period, they still end up with more ailments in the future since their health does not improve. Similar to health, survival (Figure 14) is higher if products are chosen randomly. In general, these results suggest that dynamic payoffs would rise through technology improvements under a choice policy that is not necessarily consistent with dynamic optimization. We turn to a discussion of welfare and alternative choice probabilities next.

6.3 Demand Externalities and the Social Planner's Problem

Results in the previous section point to a choice externality whereby individually rational optimal choices slow the path of technological progress. This might mean that there is a coordination failure. Here, we ask whether the social planner can raise welfare by imposing an alternative choice probability that speeds the progress or that changes the direction of innovation.

[To Be Added]

7 Conclusion

[To Be Added]

References

Acemoglu, Daron and Joshua Linn, "Market Size in Innovation: Theory and Evidence from the Pharmaceutical Industry," *Journal of Political Economy*, 2004, *119* (23), 1049–90.

Allcott, Hunt, Rebecca Diamon, and Jean-Pierre Dubé, "The Geography of Poverty and Nutrition: Food Deserts and Food Choices Across the United States," Mimeo, New York University 2015.

Bhaskaran, K, O Hamouda, M Sannes, and et al, "Changes in the Risk of Death after HIV Seroconversion Compared with Mortality in the General Population," *Journal of the American Medical Association*, 2008, *300* (1), 51–59.

Bhattacharya, Jay and Mikko Packalen, "The Other Ex Ante Moral Hazard in Health," *Journal of Health Economics*, 2012, *31* (1), 135–146.

Bolton, Patrick and Christopher Harris, "Strategic Experimentation," *Econometrica*, 1999, *67* (2), 349–374.

Chan, T.Y. and B.H. Hamilton, "Learning, Private Information, and the Economic Evaluation of Randomized Experiments," *Journal of Political Economy*, 2006, *114* (6), 997–1040.

Crawford, G.S. and M. Shum, "Uncertainty and Learning in Pharmaceutical Demand," *Econometrica*, 2005, pp. 1137–1173.

Dranove, David, Craig Garthwaite, and Manuel Hermosilla, "Pharmaceutical Profits and the Social Value of Innovation," NBER Working Paper 2014.

Duda, R.O. and P.E. Hart, *Pattern Classification and Scene Analysis*, Wiley, 1973.

Erdem, Tülin and Michael P Keane, "Decision-Making under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets," *Marketing Science*, 1996, *15* (1), 1–20.

Finkelstein, Amy et al., "Static and Dynamic Effects of Health Policy: Evidence from the Vaccine Industry," *Quarterly Journal of Economics*, 2004, *119* (2), 527–564.

Goettler, Ronald L. and Brett Gordon, "Does AMD Spur Intel to innovate More?," *Journal of Political Economy*, 2011, *119* (6), 1141–1200.

Gowrisankaran, Gautam and Marc Rysman, "Dynamics of Consumer Demand for New Durable Goods," *Journal of Political Economy*, 2012, *120* (6), 1173–1219.

Hicks, John R, *The Theory of Wages*, London: Macmillan, 1932.

Hotz, V Joseph and Robert A Miller, "Conditional Choice Probabilities and the Estimation of Dynamic Models," *The Review of Economic Studies*, 1993, *60* (3), 497–529.

_ , _ , Seth Sanders, and Jeffrey Smith, "A Simulation Estimator for Dynamic Models of Discrete Choice," *The Review of Economic Studies*, 1994, *61* (2), 265–289.

Lancaster, K.J., "A New Approach to Consumer Theory," *The Journal of Political Economy*, 1966, *74* (2), 132–157.

Papageorge, Nicholas W., "Why Medical Innovation Health Is Valuable: Health, Human Capital and the Labor Market," Forthcoming, Quantitative Economics 2016.

Petrin, A., "Quantifying the Benefits of New Products: The Case of the Minivan," *Journal of Political Economy*, 2004, *110* (4), 705–729.

Rosen, S., "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition," *The Journal of Political Economy*, 1974, *82* (1), 34–55.

Scherer, Frederic M, "Demand-Pull and Technological Invention: Schmookler Revisted," *The Journal of Industrial Economics*, 1982, pp. 225–237.

Schmookler, Jacob, *Invention and Economic Growth* 1966.

Stigler, G.J., "The Cost of Subsistence," *Journal of Farm Economics*, 1945, *27* (2), 303–314.

8 Figures and Tables

Table 1: Summary Statistics: Subjects. Visit 14-47 (1990-2007)

	Analytic Sample	Obs	Full Sample	Obs
Black	0.22	1719	0.22	2420
Hispanic	0.09	1719	0.1	2420
White	0.68	1719	0.61	2420
High School	0.14	1719	0.19	2402
Some College	0.29	1719	0.3	2402
College	0.34	1719	0.31	2402
Grad	0.23	1719	0.20	2402
Died	0.40	1719	0.42	2420
Died Conditional	0.20	1185	0.24	1782
Ever Market Product	0.83	1719	0.83	2121
Ever Trial Product	0.24	1719	0.25	2121
Ever Work	0.74	1719	0.72	2304
Ever Not Work	0.68	1719	0.75	2304
Age 1991	36.04	1719	34.86	2406
	[8.72]		[9.59]	

Notes: Standard deviation in square brackets. Data for unique individuals. *Ever Market Product* stands for ever consumed a market product during the period from visit 14 to visit 47. Similar definition holds for *Ever Trial Product*. *Died Conditional* is the proportion of individuals who died conditional on surviving until year 1995.

Table 2: Summary Statistics: Subjects-Visits. Visits 14-47 (1990-2007)

	Analytic Sample	Pre Haart	Post Haart
Obs	16851	6972	9879
Ailments	0.43	0.45	0.41
Market Product	0.65	0.49	0.76
Trial Product	0.07	0.09	0.05
Work	0.63	0.70	0.58
Age	44.48	40.89	47.01
	[8.03]	[6.99]	[7.75]
CD4	475	407	524
	[297]	[298]	[287]
Gross Income	17567	19036	16531
	[8787]	[8733]	[8677]
Out-of-pocket Pay	266	179	327
	[706]	[598]	[767]

Notes: Standard deviation in square brackets. Income and Out-of-pocket are semestral and measured in real dollars of 2000. Pre HAART era corresponds to visit <= 24 or (roughly) year <= 1995.

Table 3: Distribution of Number of New Products, F_N

	coef.	se
μ		
$MaxChange_{t-1}$	0.432	0.246
$TrialsShare_{t-1}$	6.177	2.462
$\ln \alpha$		
$Constant$	-0.206	0.451
$MaxChange_{t-1}$	-1.019	0.626

Notes: Model specified in (22). The variable $MaxChange_{t-1}$ measures the distance between the previous period's new products and the previous period's centroid. It captures the relatively higher number of new products that follow the appearance of better innovations. The variable $TrialsShare_{t-1}$ is the share of individuals going into a trial the previous period. According to the model in (22), $E[New_t] = \mu$ and $Var[New_t] = \mu(1 + \alpha\mu)$.

Table 4: Utility Parameters, y_{it}

parameter	variable	coef.	se
α_{4w}	$1\{cluster\} \cdot white$	-1.385	0.206
α_{4b}	$1\{cluster\} \cdot black$	-1.868	0.210
α_{4l}	$1\{cluster\} \cdot hispanic$	-1.075	0.835
α_{4a}	$1\{cluster\} \cdot a_{it-1}$	0.003	0.005
α_{4h}	$1\{cluster\} \cdot h_{it-1}/10^3$	-3.385	0.134
α_{5w}	$1\{trial\} \cdot white$	-2.678	0.168
α_{5b}	$1\{trial\} \cdot black$	-3.755	0.170
α_{5l}	$1\{trial\} \cdot hispanic$	-2.902	0.354
α_{5a}	$1\{trial\} \cdot a_{it-1}$	0.051	0.003
α_{5h}	$1\{trial\} \cdot h_{it-1}/10^3$	-1.702	0.082
α_{6w}	$1\{stay\} \cdot white$	0.525	0.157
α_{6b}	$1\{stay\} \cdot black$	0.396	0.159
α_{6l}	$1\{stay\} \cdot hispanic$	0.480	0.674
α_{6a}	$1\{stay\} \cdot a_{it-1}$	0.019	0.003
α_{6h}	$1\{stay\} \cdot h_{it-1}/10^3$	1.048	0.101
α_x	x_{it}	0.522	0.292
α_{xp}	$x_{it} \cdot 1\{product\}$	-3.575	0.226
α_m	$m_{it} - o_{it}$	0.141	0.023

Notes: Estimation of equation (9). Discount factor $\beta = .8$. $1\{cluster\}$ indicates whether the individual chose one of the three clusters of products available. $1\{product\}$ indicates whether the individual consumes a product in t, $1\{product\} = 1\{cluster\} + 1\{stay\} + 1\{trial\}$.

Table 5: Gross Income, m_{it}

variable	coef.	se
h_{it-1}	0.018	0.004
$h^2_{it-1}/10^3$	-0.064	0.019
$h^3_{it-1}/10^7$	1.138	0.381
$h^4_{it-1}/10^{10}$	-1.030	0.381
$h^5_{it-1}/10^{14}$	4.854	1.950
$h^6_{it-1}/10^{18}$	-11.270	4.850
$h^7_{it-1}/10^{20}$	0.101	0.046
a_{it-1}	0.482	0.114
a^2_{it-1}	-0.006	0.001
black	-5.534	0.366
hispanic	-4.167	0.570
some college	2.497	0.442
college	5.812	0.457
more than college	8.203	0.500
l_{it}	5.738	0.220
x_{it}	0.207	0.084
constant	-2.095	2.620

Notes: Estimation of equation (11). Random effects regression of gross-income on covariates. m_{it} is measured in thousands of real dollars of 2000. Health is given by the CD4 count measured in hundreds of cells per microliter.

Table 6: Tobit Model for Out-of-pocket Payments, o_{it}

variable	coef.	se
h_{it-1}	-0.002	0.000
$h_{it-1}^2/10^3$	0.009	0.002
$h_{it-1}^3/10^7$	-0.133	0.029
$h_{it-1}^4/10^{10}$	0.090	0.022
$h_{it-1}^5/10^{14}$	-0.266	0.071
$h_{it-1}^6/10^{18}$	0.279	0.083
a_{it-1}	0.037	0.007
a_{it-1}^2	0.000	0.000
black	-0.240	0.021
hispanic	-0.119	0.025
some college	0.169	0.026
college	0.318	0.033
more than college	0.336	0.030
market product	0.429	0.026
trial product	0.313	0.043
l_{it}	0.105	0.016
x_{it}	-0.122	0.017
constant	-1.459	0.182
σ^o	0.862	0.066

Notes: Estimation of equation (12). *market product* $= \sum_{k=1}^{4} d_{kit}$. o_{it} is measured on thousands of real dollars of 2000. Health is given by the CD4 count measured in hundreds of cells per microliter.

Table 7: Logit Model for Labor Supply, l_{it}

variable	coef.	se
h_{it-1}	0.009	0.001
$h_{it-1}^2/10^3$	-0.013	0.002
$h_{it-1}^3/10^7$	0.075	0.023
$h_{it-1}^4/10^{10}$	-0.013	0.007
a_{it-1}	0.102	0.032
a_{it-1}^2	-0.001	0.000
black	-0.168	0.073
hispanic	-0.040	0.125
some college	0.312	0.105
college	0.537	0.103
more than college	0.613	0.108
l_{it-1}	4.458	0.056
constant	-5.914	0.742

Notes: Estimation of equation (13). Health is given by the CD4 count measured in hundreds of cells per microliter.

Table 8: Logit model for Death, $1 - S_{it+1}$

variable	coef.	se
h_{it-1}	-0.028	0.003
$h_{it-1}^2/10^3$	0.079	0.015
$h_{it-1}^3/10^7$	-1.104	0.292
$h_{it-1}^4/10^{10}$	0.704	0.220
$h_{it-1}^5/10^{14}$	-1.610	0.561
a_{it-1}	-0.116	0.058
a_{it-1}^2	0.002	0.001
black	-0.509	0.199
hispanic	0.034	0.235
some college	0.060	0.185
college	-0.353	0.185
more than college	-0.512	0.207
x_{it}	-1.140	0.159
constant	1.682	1.358

Notes: Estimation of equation (17). Health is given by the CD4 count measured in hundreds of cells per microliter.

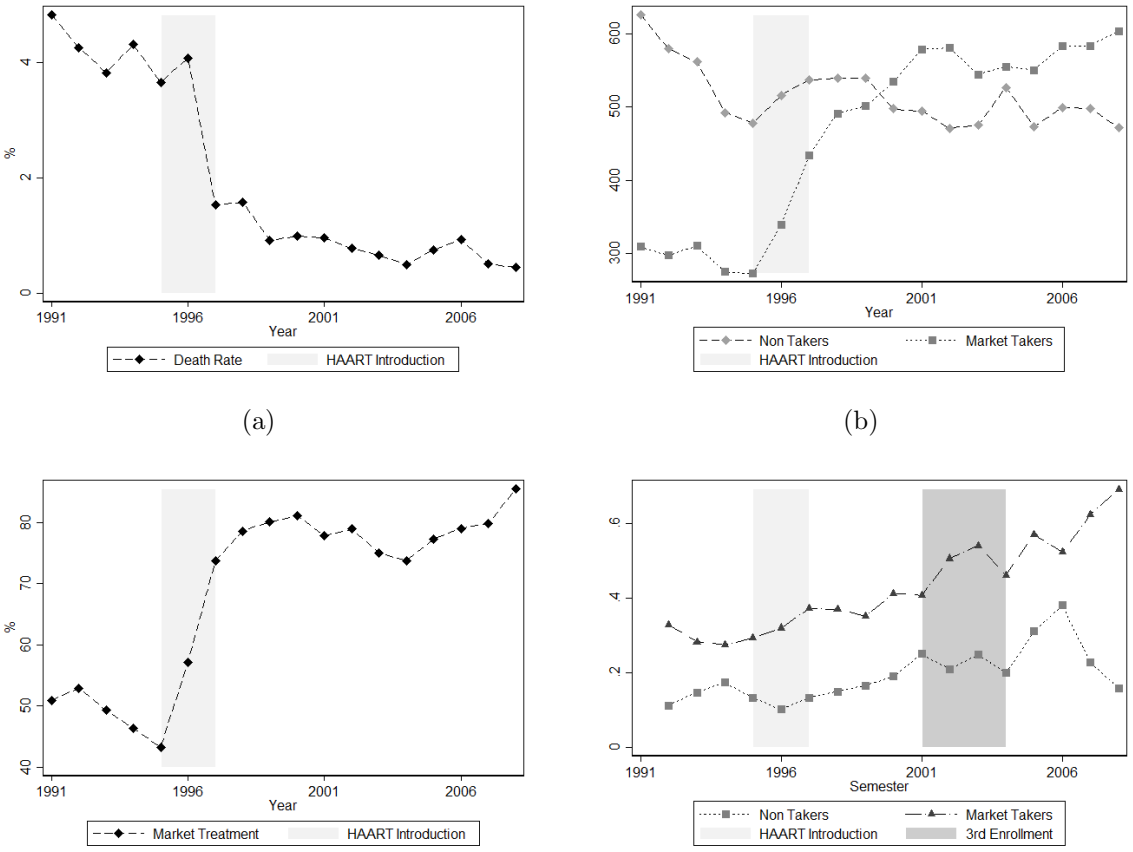


Figure 1: Summary trends over time. Panel 1(a): Death probability over time. Panel 1(b): Mean CD4 over time by consumption status. Panel 1(c): Market treatment consumption over time. Panel 1(d): Mean out-of-pocket payments over time.

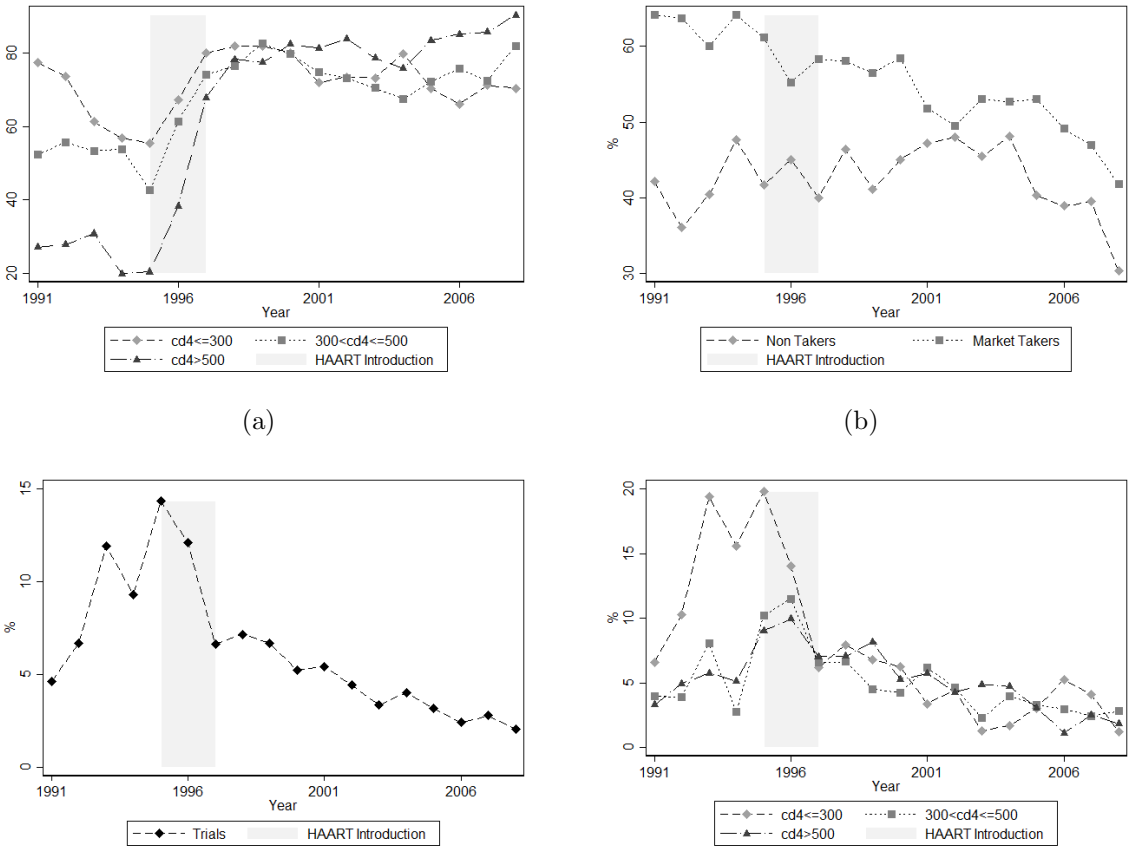


Figure 2: Summary trends over time. Panel 2(a): Treatment usage by CD4 over time. Panel 2(b): Mean ailments over time. Panel 2(c): Trial participation over time. Panel 2(d): Trial participation by CD4 over time.

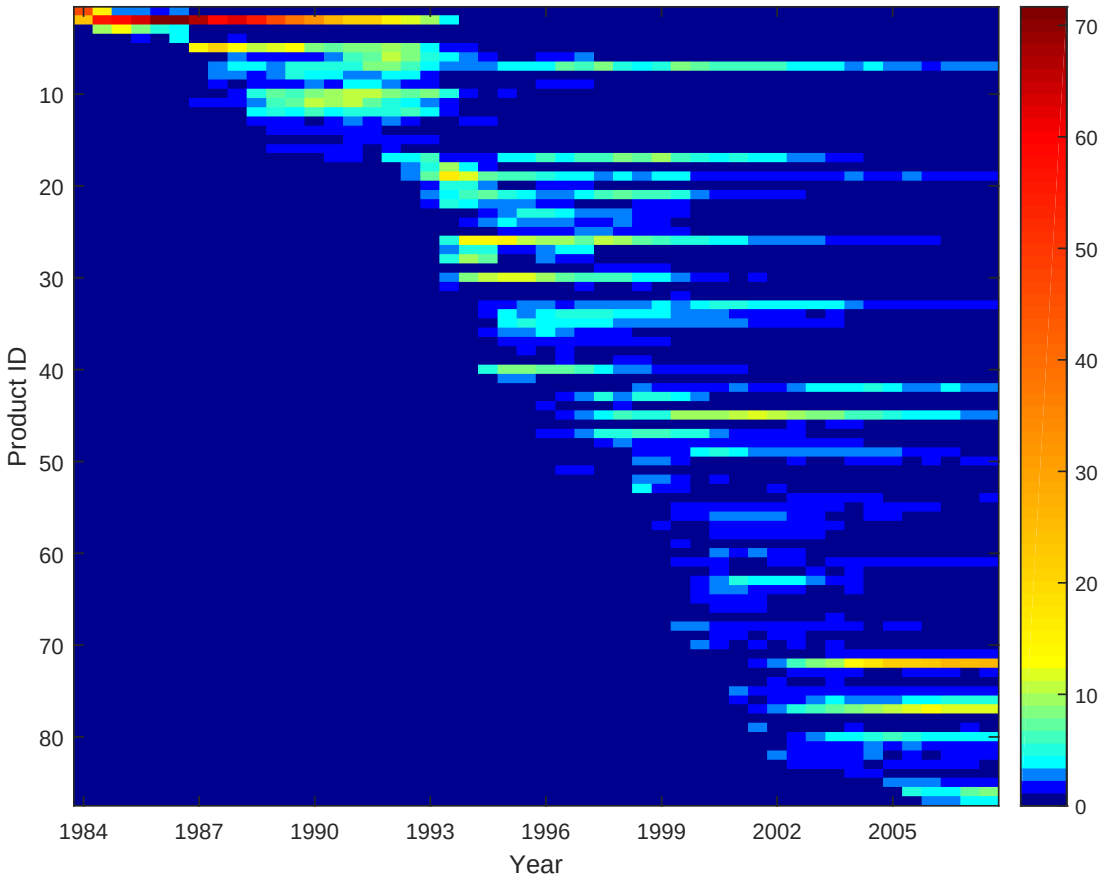


Figure 3: Diffusion of Products Over Time

Notes: Each id—or row—represents a product. Color indicates the share of the market that the product captures. Shares are conditional on consuming a product.

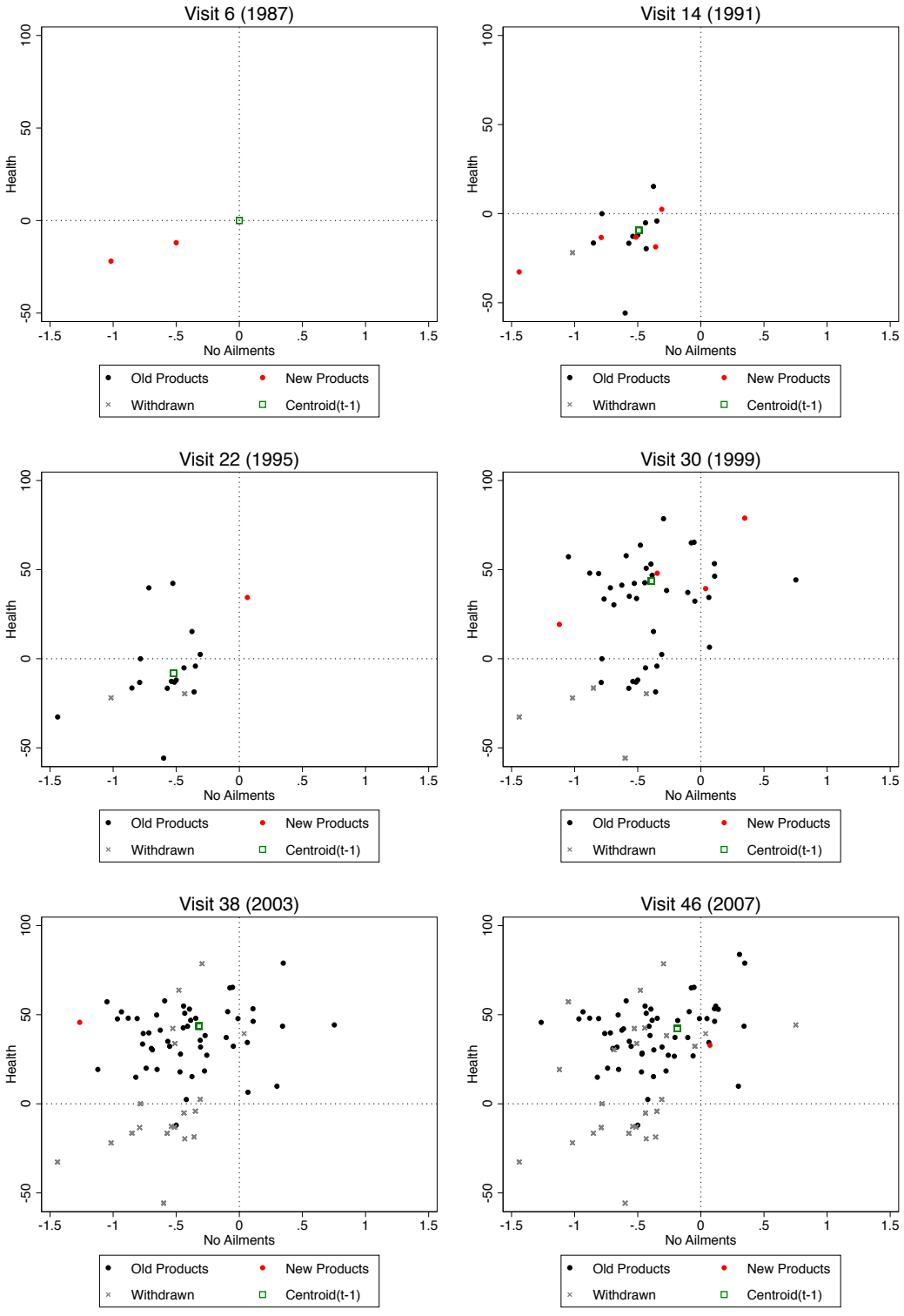


Figure 4: Treatment Evolution

Notes: Figure shows the evolution of the market of HIV treatments. Every product is identified by its characteristics, θ, in the health and no-ailments axises.

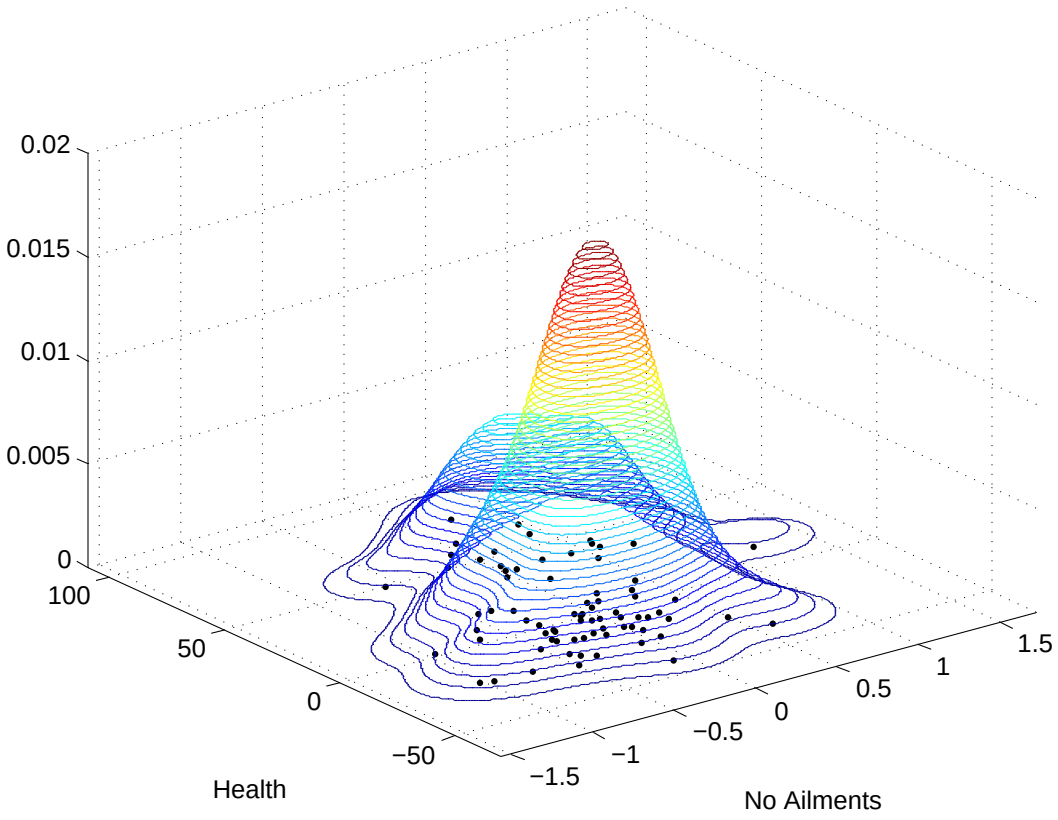


Figure 5: Smoothed Kernel of Empirical Innovations Distribution

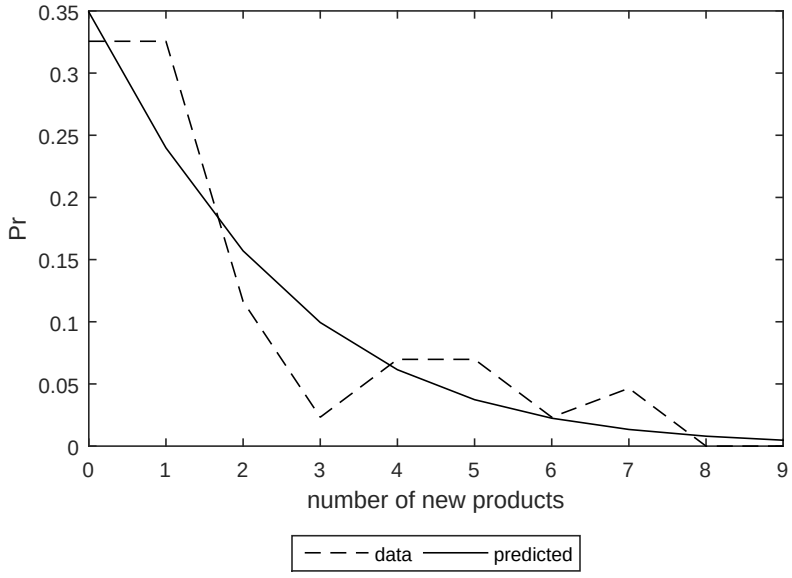


Figure 6: Distribution of Number of New Products

Notes: Model specified in (22). Figure shows the empirical distribution of new products and the average over time of the predicted probabilities using the estimated parameters in Table 3.

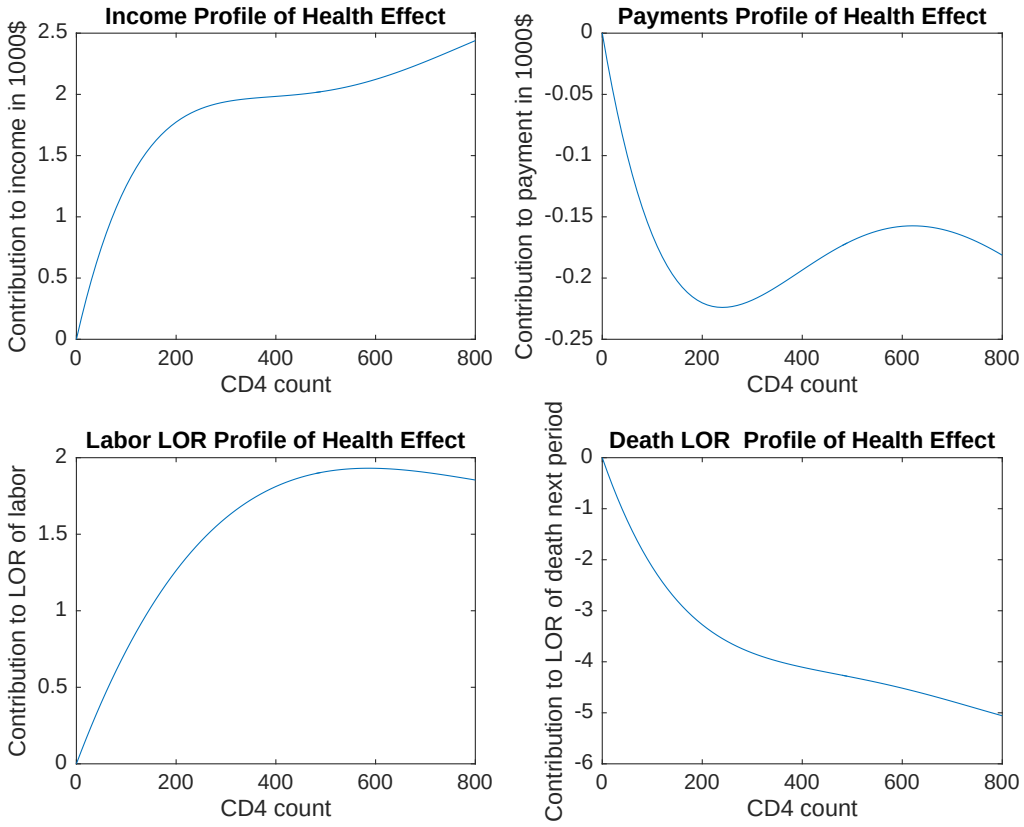


Figure 7: Health Effect on Outcomes

Notes: Semestral income measured in thousands of dollars of 2000. CD4 Count measured in hundreds of cells per microliter. LOR stands for log odds ratio.

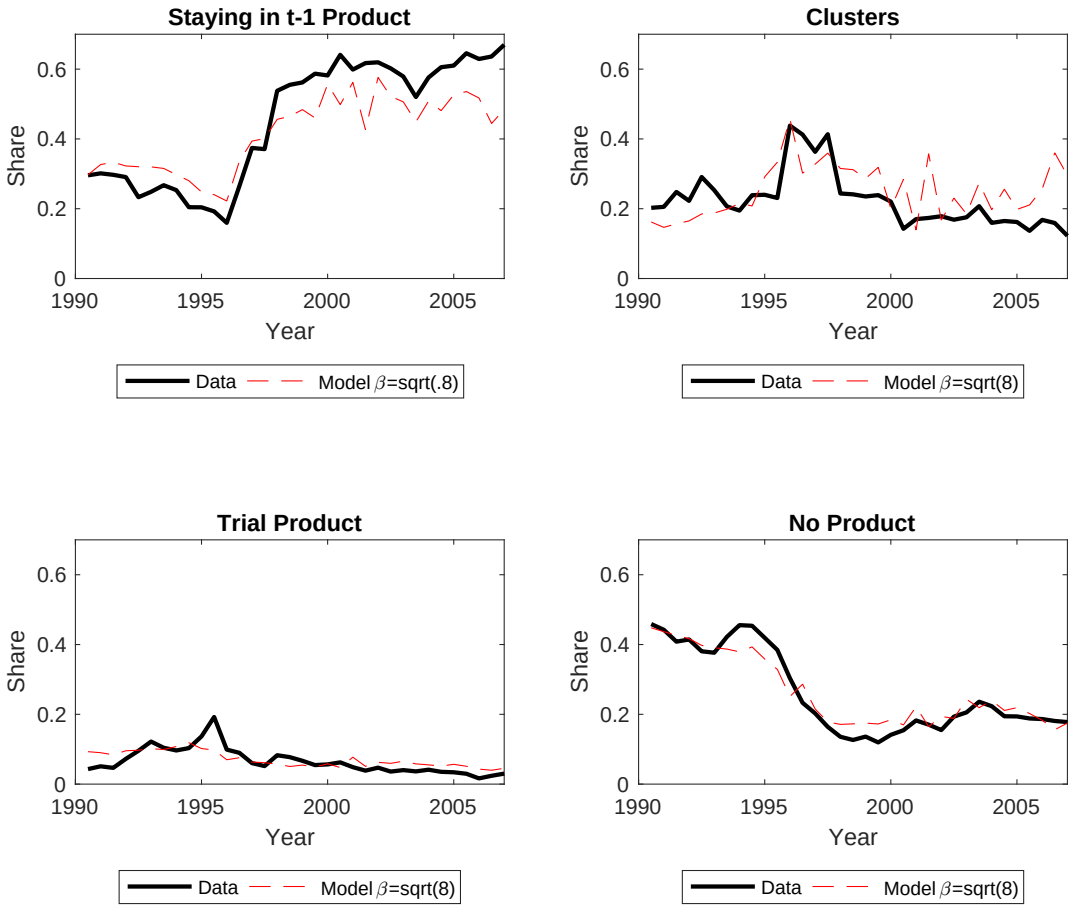


Figure 8: Goodness of Fit Figures

Notes: Simulated and empirical choice rates over time.

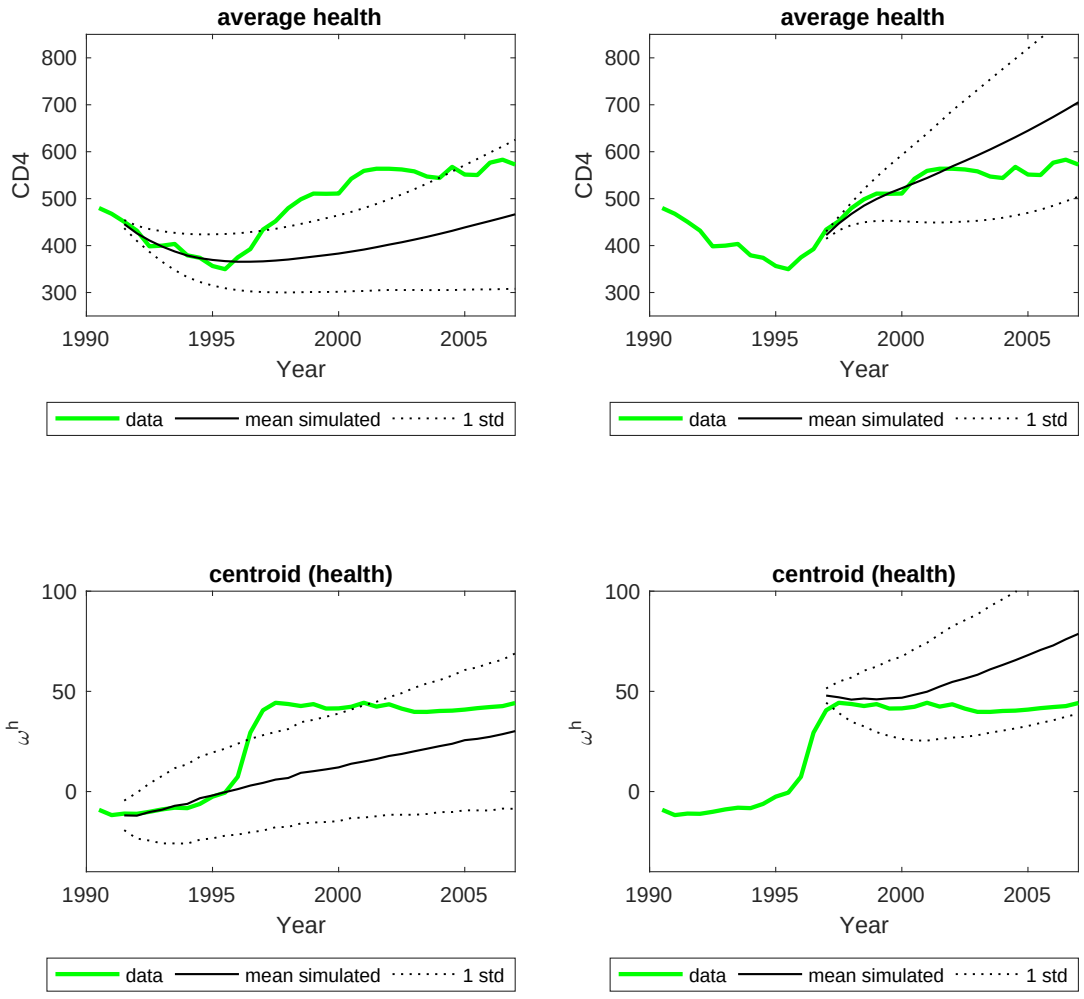


Figure 9: Distribution of Technology Paths: Health

Notes: statistics computed over 1000 simulated paths conditional on the state of the world at 1991 and 1997.

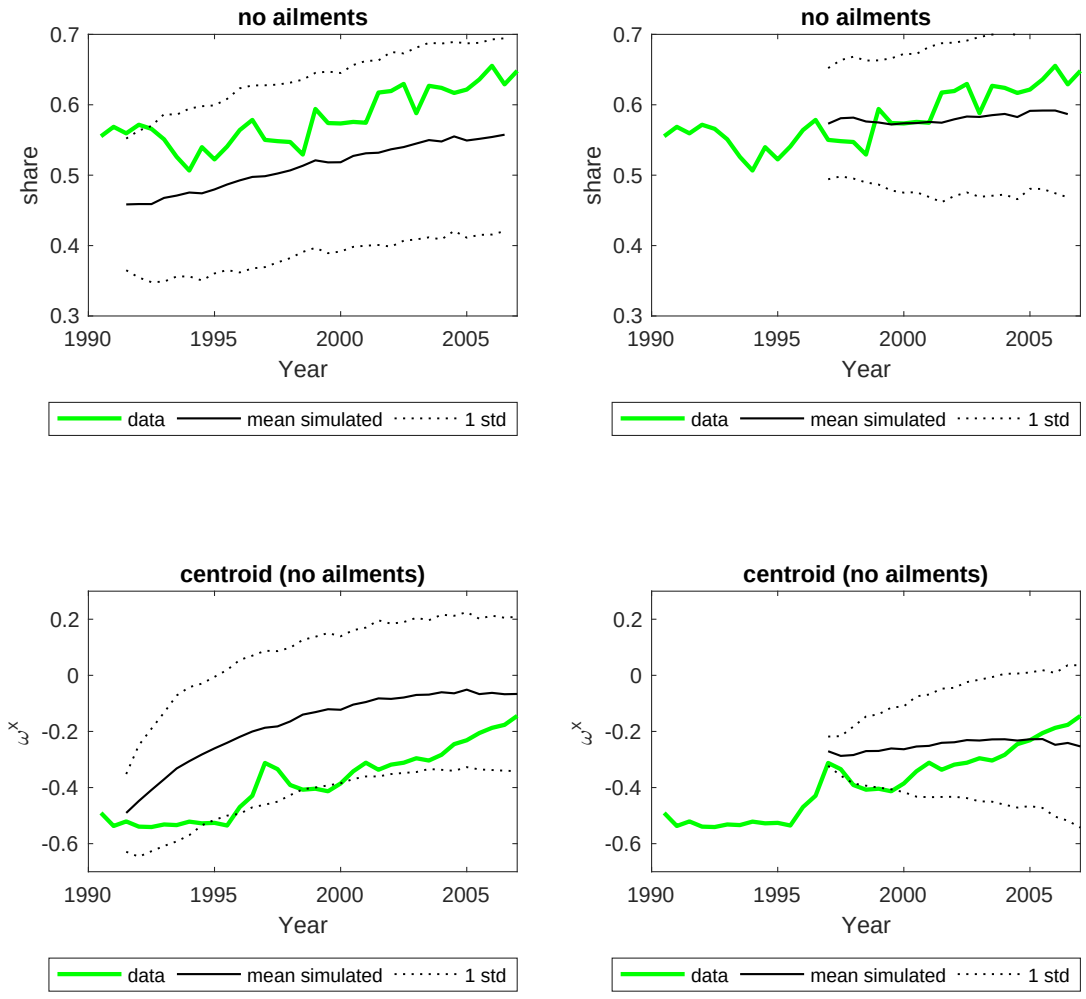


Figure 10: Distribution of Technology Paths: No Ailments

Notes: statistics computed over 1000 simulated paths conditional on the state of the world at 1991 and 1997.

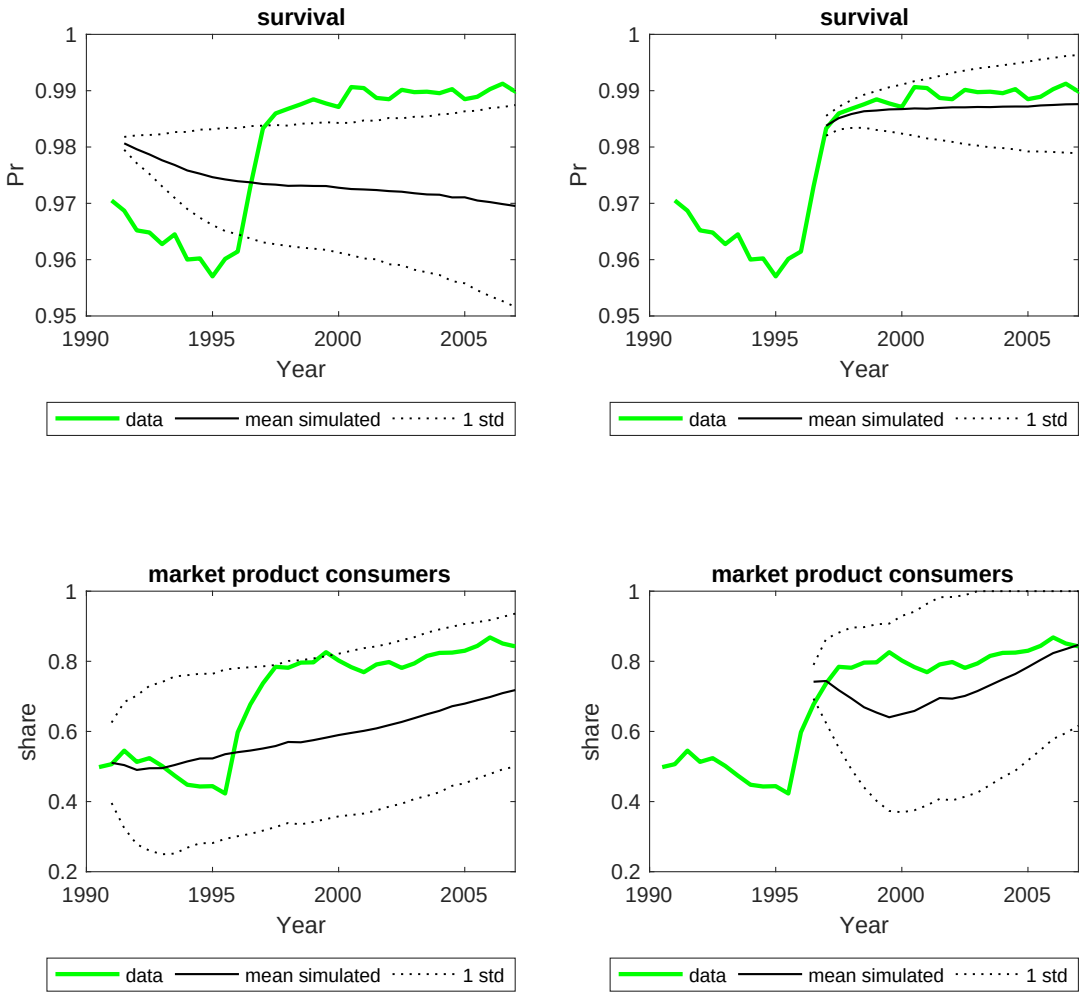


Figure 11: Distribution of Technology Paths: Survival and Product Consumption

Notes: statistics computed over 1000 simulated paths conditional on the state of the world at 1991 and 1997.

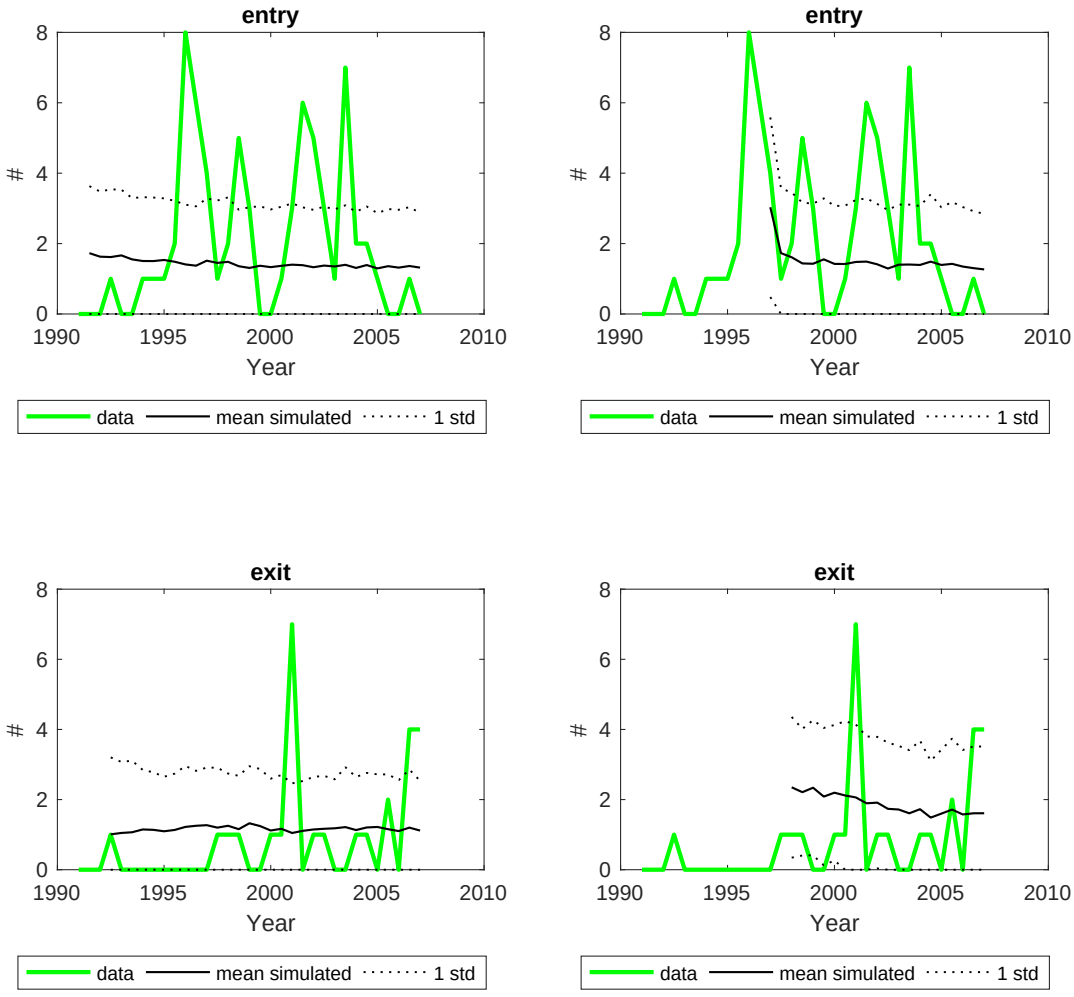


Figure 12: Distribution of Technology Paths: Entry and Exit

Notes: statistics computed over 1000 simulated paths conditional on the state of the world at 1991 and 1997.

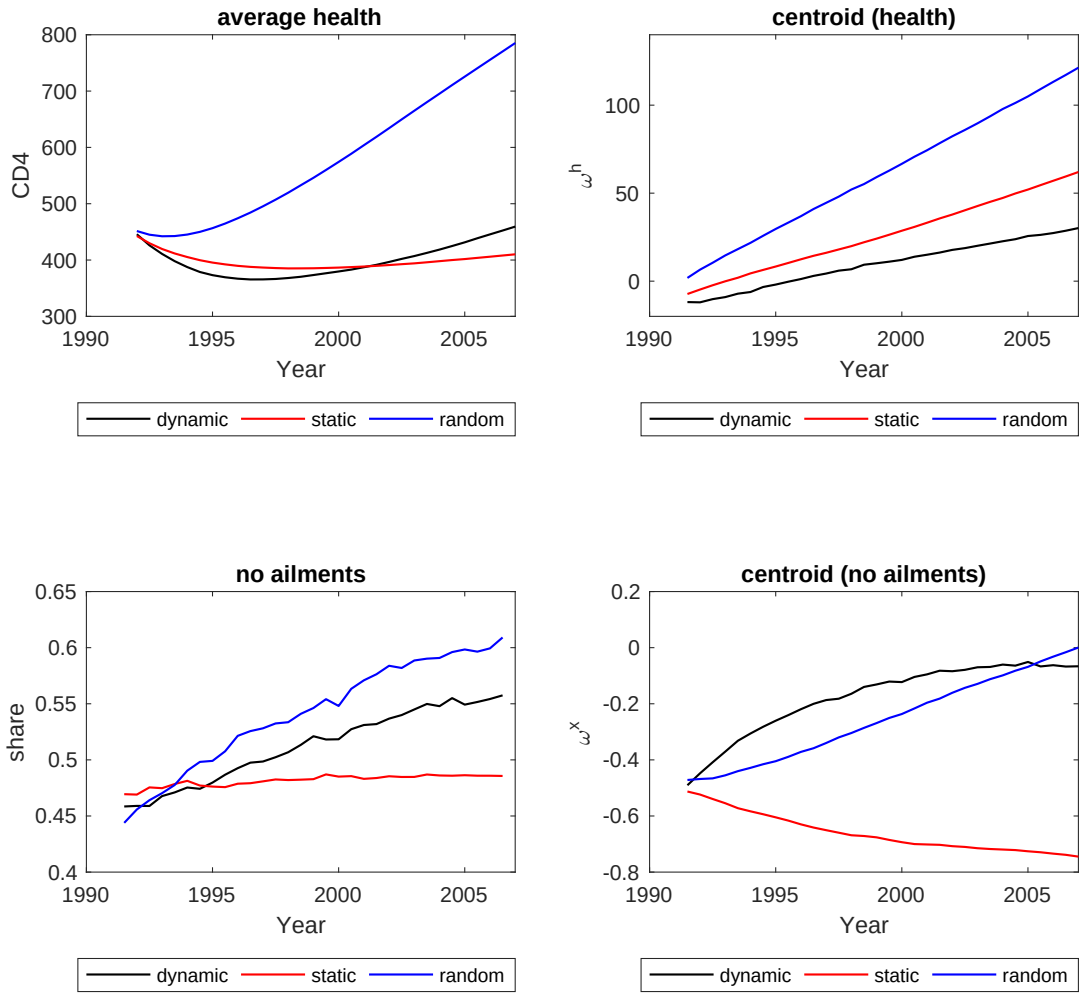


Figure 13: Alternative Choice Regimes: Health and No Ailments

Notes: Alternative choice regimes are: (i) optimal dynamic choice, (ii) static optimal choice, and (iii) random choice. Mean over 1000 simulated paths of the relevant statistic conditional on the state of the world at 1991 and 1997.

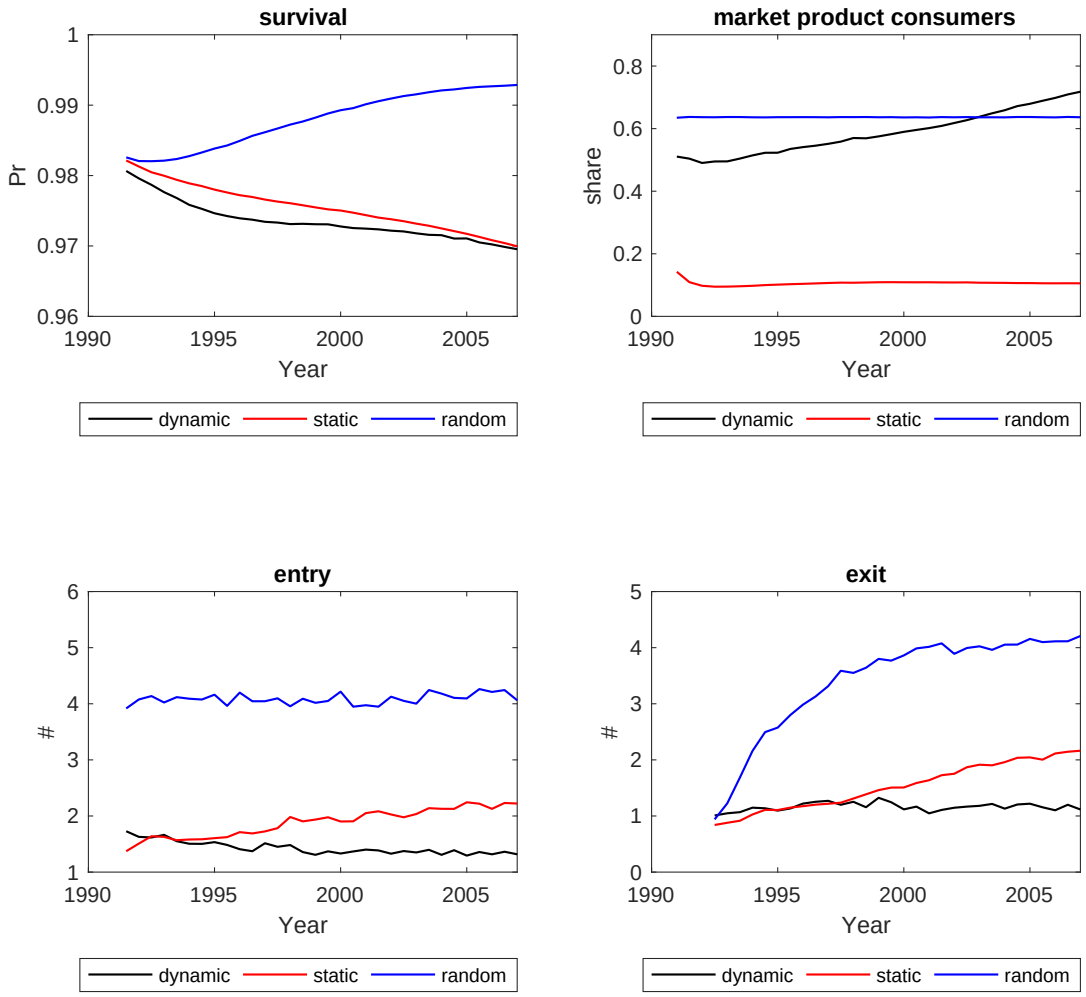


Figure 14: Alternative Choice Regimes: Survival, Consumption, Entry, and Exit

Notes: Alternative choice regimes are: (i) optimal dynamic choice, (ii) static optimal choice, and (iii) random choice. Mean over 1000 simulated paths of the relevant statistic conditional on the state of the world at 1991 and 1997.

A Data Appendix

Beginning in 1984, the Multi-Center AIDS Cohort Study (MACS) started gathering information regarding natural and treated histories of HIV infection in homosexual and bisexual men. The study is conducted in Baltimore, Chicago, Pittsburgh and Los Angeles. At each semi-annual visit, data are collected on: demographics, psychosocial characteristics, sexual behavior, and specially important for our purposes, antiretroviral (AV henceforth) drugs consumption and trial participation. In addition, blood tests are administered to measure health status and serostatus (whether the individual is HIV+). Data collection started with 4,954 men enrolled. Two more enrollments have taken place: one in 1987-1991 (668 additional men) and another in 2001-2003 (1,350 additional men). We only use data from the first two enrollments.

A.1 Main Variables

Health (h_{it-1}): At every visit individuals go through a physical examination in which several health measurements are taken. As our measure of underlying health status, we use the CD4 count obtained from a blood sample. "CD4 is a glycoprotein found on the surface of immune cells [...]. If CD4 cells become depleted, for example in untreated HIV infection, or following immune suppression prior to a transplant, the body is left vulnerable to a wide range of infections that it would otherwise have been able to fight. [...] Normal blood values are usually expressed as the number of cells per microliter (or cubic millimeter, mm^3) of blood, with normal values for CD4 cells being 500-1200 cells/mm" (Wikipedia). We denote as h_{it-1} the CD4 count at of the individual at the start of period t.

Labor supply (l_{it-1}): Whether the individual was working full time (35 hours or more) in visit t.

Income (m_{jit}): Starting at visit 14, individuals answer the following question: "Which of the following categories describes your annual individual gross income before taxes"? For visit 14, categories are: less than 10000, 10000-19999, 20000-29999, 30000-39999, 40000-49999, 50000-59999, 60000-69999, 70000 or more, Doesn't wish to answer. For visits 15 to 35, categories are: less than 10000, 10000-19999, 20000-29999, 30000-39999, 40000-49999, 50000 or more, Doesn't wish to answer. For visits 36 to 41, categories are: less than 10000, 10000-19999, 20000-29999, 30000-39999, 40000-49999, 50000-99999, 60000 or more, Doesn't wish to answer.

We censor all periods at 50000 or more to obtain a uniform question over time. Then we assign the middle point to individuals in the bracket. For the highest bracket we assign

the upper limit (50000). In our model gross income is divided by two since the survey asks about annual income. Gross income as well as out-of-pocket payments (below) are in constant dollars of 2000.

Out-of-pocket payments (o_{jit}): Starting from visit 14, individuals are given the following direction "Please estimate the TOTAL out-of-pocket expenses that you or other personal sources (your lover, family or friends) paid for prescription medications since your last visit."[20] As opposed to the gross income question, this question is open so values are not categorized.

Ailments (x_{jit}): Starting from at visit 4, individuals are asked about physical symptoms. Even though other ailments are recorded, we focus on unusual bruises lasting at least two weeks, unintentional weight loss of at least 10 pounds, fatigue, diarrhea, fever, night sweats, and tender/enlarged glands. The last 5 ailments must be felt for at least 3 days.

Even though individuals are asked explicitly about side effects starting from visit 13, we choose not to use such data because it is less consistent and, more importantly, because we do not think individuals are able to differentiate correctly between side effects and symptoms. Therefore, in our model x_{it} takes the value of 1 if individual reports having any of the problems mentioned above.

Race (b_i): Individuals in the sample are either white, black or hispanic.

Age (a_{it}): Age of the individual at the beginning of period t.

A.2 Products and Product Components

At every visit after visit 6, individuals are asked whether they took any medication to fight AIDS. Starting from visit 13, as the number of medications becoming available for HIV exploded, separate surveys were administered for antiretroviral drugs (ARVs) and non antiretroviral drugs (NARVs). We focus on ARVs since these are the drugs used to treat HIV infection. Further, since our analysis includes estimating the health and ailments of people using different drugs, we focus on observations where individuals have reported a treatment along with h_{it}, h_{it-1}, and x_{it}.

Individuals are asked to name specifically which drugs they took as well as whether or not they took the drug as part of a research study (the exact wording of the question regarding research studies changes slightly over time). Some of the reported drugs are themselves coded as trials; we regard these instances as individuals participating in trials (see Table S1). If at individual i at period t is consuming one of his drugs as part of a trial we regard

[20] Wording changes slightly in visits 14 and 15.

individual i as consuming a trial product at period t.

Next, we define market products as treatments with no components consumed in trial. Given that the sum of effects of individual drugs is not equal to the effect of a treatment formed by the sum of the drugs, the relevant market products consumed in our data correspond to combinations of components. For instance a product is AZT and another is AZT plus 3TC plus ddI. Table S2 describes the individual components of market products. Some components, listed separately in Table S3, are in fact fixed-dose combinations of other components. In our sample, if individual i is consuming the fixed-dose combination $(A+B)$ and individual i' is consuming components A and B, we assign consumers i and i' to the same treatment. One of the coded components in the data corresponds to "other ARVs". We add all uncoded components (96 instances) to "other ARVs" which results in 158 instances of "other ARVs". Finally, we treat α and β Interferons (177 instances and 33 instances, respectively) as one single component.

Our definition of market products, as combinations of drug components, generates 1835 different market treatments. We reduce the number of market products using the following algorithm:

1. We select our core market products as those treatments that overall have more than 40 instances.[21] We acknowledge that our definition of core treatments is biased against treatments appearing near the end of the time period studied. We address this issue by excluding the last 4 periods of data. Our core treatments are listed in Table S4 which shows that there are 70 core products overall and they have at most five components. Out of 20142 subject-visit observations of consumers taking market products, 13767 are covered by core treatments and 6375 correspond to non-core treatments.

2. Second, we assign non-core treatments to core treatments in the following fashion. Each step is used sequentially to assign remaining non core treatments that were not assigned in previous steps.

   (a) Assignment of Non-core: Non core treatment A is assigned to core treatment B if B is the core treatment with the highest number of components that is contained by A. This procedure yields both non-unique assignments or null assignments. Of the remaining 6375 subject-visit observations of non core treatments, 2963 are assigned uniquely in this step. This means that we are left with 3412 subject-visit observations with non core treatments, 1647 that are assigned to multiple core treatments and 1765 that are not assigned to any core treatment.

[21] We can change this to a different number and main results remain robust.

(b) Assignment of Multiple Assignments:

i. First, we use the past history of the individual. If at period t individual i is consuming non core treatment W that was assigned to both core treatments A and B in previous steps, and he was observed consuming core treatment A in period $t-1$, then his treatment at t is assigned uniquely as A. We repeat this procedure until no further gains are obtained. Out of the remaining 1647 subject-visit observations of non core treatments with multiple assignments, 428 are assigned uniquely in this step.

ii. Second, we use the future history of the individual. If at period t individual i is consuming non core treatment W that was assigned to both core treatments A and B in previous steps, and he was observed consuming core treatment B in period $t+1$, then his treatment at t is assigned uniquely as B. We repeat this procedure until no further gains are obtained. Out of the remaining 1219 subject-visit observations of non core treatments with multiple assignments, 274 are assigned uniquely in this step.

iii. Third, we assign the reaming 945 subject-visit observations of non core treatments with multiple assignments using the core treatment with the highest share at t: if at period t individual i is consuming non core treatment W that was assigned to both core treatments A and B in previous steps, and treatment A's market share at t is greater than B's, his treatment at t is assigned uniquely as A. This final step assigns uniquely the remaining 945 observations.

(c) Next, we regard all 1765 not assigned treatment observations as "fringe" treatments since they do not contain any core treatment. We aggregate them in the following fashion. We aggregate all fringe treatments that appear at period t and assign to this "cohort" fringe treatment, all users consuming this product over time. Similarly as we do with core treatments, we only consider fringe cohort treatments that have at least 40 users. This reduces the number of observations by 345 (which represents 1.6% of the number of observations of treatment consumers). This aggregations leads to 17 fringe cohort treatments that we will treat in the same way we treat core treatments: as innovations from the trials distribution. This amounts to a total of 87 treatments over all. From this point on fringe treatments are included in the denomination of core treatments.

3. We have specified that a treatment gets withdrawn from the market if it has zero share for $X = 2$ consecutive periods. However, in the data, a treatment may have zero share

for $Y > X$ periods and then reappear again. 78 out of 87 core treatments have unique spells; we regard the remaining treatments with multiple spells as measurement error and follow the next procedure to ensure that treatments have one single spell from entry to exit. Consider a core treatment with multiple spells B.

(a) We identify all spells that treatment B has in the data.

(b) Among treatment B's spells, we select the spell that contains the period in which treatment B's share was the highest. We drop all observations of market consumers of treatment B that are not in this spell.

(c) We follow the same steps for all 9 core treatments with multiple spells. Out of 19797 ($20142 - 345$) subject-visit observations of consumers taking market products, this smoothing procedure drops 42 observations leaving 19755 subject-visit observations of consumers taking market products.

As evidence of the relevance of the spells selected by this procedure we compute the difference between the maximum share in the selected spell and the maximum share in each of the other spells, as a percentage of the maximum share in the other spell. The mean value of this measure is 2401, which suggests that the maximum share in the selected spell is on average about 24 times larger that the maximum share in other spells. We also try the following criteria: (i) selecting the spell with the highest average share and (ii) selecting the spell with the highest sum of shares. All criteria result in closely similar entry and exit dates so we stick to the maximum-share criteria.

Appendix Table S1: Trial Components

Name	Observations
AZT/ddI Blinded Trial	91
AZT/ddC Blinded Trial	69
ddI/ddC Blinded Trial	6
AZT/ddI/ddC Blinded Trial	31
AZT/d4T trial	4
AZT/3-TC Blinded Trial	23
AZT/ddI/protease inhibitor Blinded Trial	1
AZT/protease inhibitor Blinded Trial	2
d4T/protease inhibitor Blinded Trial	1
AZT/3-TC/protease inhibitor Blinded Trial	1
Combivir/Trizivir Blinded Trial	5
Trizivir + Sustiva/Combivir + Sustiva Blinded Trial	3
Generic AIDS Vaccine Trial	1

Appendix Table S2: Chemical Formulae of Product Components

Component	Chemical Formula	Observations
Isoprinosine	$C_{52}H_{78}N_{10}O_{17}$	87
Ribavirin	$C_8H_{12}N_4O_5$	62
Interferons (α and β)		210
Foscarnet	CH_3O_5P	92
AZT	$C_{10}H_{13}N_5O_4$	7436
ddC	$C_9H_{13}N_3O_3$	1123
AL-721 egg lecithin		147
Dextran-Sulfate	$H(C_6H_{10}O_5)_xOH$	65
Acyclovir	$C_8H_{11}N_5O_3$	2550
ddI	$C_{10}H_{12}N_4O_3$	3069
d4T	$C_{10}H_{12}N_2O_4$	3807
Nevirapine	$C_{15}H_{14}N_4O$	2210
Delavirdine	$C_{22}H_{28}N_6O_3S$	176
3TC	$C_8H_{11}N_3O_3S$	5250
Saquinavir	$C_{38}H_{50}N_6O_5$	1279
Ritonavir	$C_{37}H_{48}N_6O_5S_2$	3230
Indinavir	$C_{36}H_{47}N_5O_4$	2255
Nelfinavir	$C_{32}H_{45}N_3O_4S$	1278
Kaletra	$C_{37}H_{48}N_4O_5$	1883
Abacavir	$C_{14}H_{18}N_6O$	1549
Agenerase	$C_{25}H_{35}N_3O_6S$	372
Efavirenz	$C_{14}H_9CIF_3NO_2$	3362
Adefovir	$C_8H_{12}N_5O_4P$	44
Enfuvirtide (T-20)	$C_{204}H_{301}N_{51}O64$	160
Tenofovir	$C_9H_{14}N_5O_4P$	2488
Emtricitabine	$C_8H_{10}FN_3O_3S$	263
Atazanavir	$C_{38}H_{52}N_6O_7$	1583
Lexiva	$C_{25}H_{36}N_3O_9PS$	418
Etravirine	$C_{20}H_{15}BrN_6O$	155
Darunavir	$C_{27}H_{37}N_3O_7S$	315
Raltegravir	$C_{20}H_{21}FN_6O_5$	384
Ampligen	Double-stranded RNA compound	25
Peptide T	$C_{35}H_{55}N_9O_{16}$	30
DTC	$C_5H_{10}NS_2Na$	10
CD4		2
Other protease		31
Vistide (cidofovir)	$C_8H_{14}N_3O_6P$	2
Tipranavir (PNU-140690)	$C_5H_{10}NS_2Na$	30
Other Avs		158

Notes: Source: Wikipedia (November, 2014)

Appendix Table S3: Combination Components

Name	Combination	Instances
Combivir	AZT + 3TC	2673
Trizivir	AZT + 3TC + Abacavir	778
Truvada	Emtricitabine + Tenofovir	1933
Epzicom	Abacavir + 3TC	724
Atripla	Efavirenz + Emtricitabine + Tenofovir	968

Appendix Table S4: Products in the Market, Entry and Exit

Treatment Id	Treatment	Haart	Entry (visit)	Exit (visit)
3	AZT	0	6	-
13	Interferons (α and/or β), AZT	0	7	23
9	AL-721 egg lecithin	0	7	15
34	AZT, Acyclovir	0	11	32
33	Acyclovir	0	11	32
47	AZT, Acyclovir, ddI	0	12	26
51	Acyclovir, ddI	0	12	32
14	AZT, ddC	0	12	35
39	AZT, ddI	0	12	41
46	ddI	0	12	-
69	AZT, ddC, Acyclovir, ddI	0	14	26
65	AZT, ddC, Acyclovir	0	14	31
67	AZT, ddC, ddI	0	14	23
63	ddC, Acyclovir	0	14	27
64	ddC	0	14	30
85	d4T	0	18	-
117	AZT, Acyclovir, 3TC	0	21	32
124	AZT, 3TC	0	22	-
146	Acyclovir, d4T, 3TC	0	23	32
161	AZT, 3TC, Saquinavir	1	24	42
157	d4T, 3TC	0	24	-
185	AZT, 3TC, Saquinavir, Ritonavir	1	25	-
164	AZT, Acyclovir, 3TC, Indinavir	1	25	32
171	Acyclovir, d4T, 3TC, Indinavir	1	25	32
169	AZT, 3TC, Ritonavir, Indinavir	1	25	45
214	d4T, 3TC, Ritonavir, Indinavir	1	25	45
254	d4T, 3TC, Saquinavir, Ritonavir	1	25	41
202	ddI , d4T, Indinavir	1	25	41
175	d4T, 3TC, Indinavir	1	25	48
165	AZT, 3TC, Indinavir	1	25	-
242	d4T, Nevirapine, 3TC	1	26	-
236	AZT, Nevirapine, 3TC	1	26	-
268	AZT, 3TC, Nelfinavir	1	26	-
377	ddI , d4T, Nelfinavir	1	26	43
292	d4T, 3TC, Nelfinavir	1	27	-
349	ddI , d4T, Nevirapine	1	27	-
311	ddI , 3TC, Nelfinavir	1	27	-
615	ddI , d4T, Efavirenz	1	29	48
644	3TC, Abacavir, Efavirenz	1	29	-
573	AZT, Nevirapine, 3TC, Abacavir	1	30	-
720	AZT, 3TC, Abacavir, Efavirenz	1	30	-

548	AZT, 3TC, Efavirenz	1	30	-
701	AZT, 3TC, Abacavir	0	30	-
532	d4T, 3TC, Efavirenz	1	30	44
581	Nevirapine, 3TC, Abacavir	1	31	-
782	d4T, 3TC, Kaletra	1	34	44
940	3TC, Kaletra, Abacavir	1	35	-
869	AZT, 3TC, Kaletra	1	35	-
987	AZT, 3TC, Kaletra, Abacavir	1	36	-
963	3TC, Abacavir, Efavirenz, Tenofovir	1	36	-
921	AZT, 3TC, Abacavir, Tenofovir	1	36	-
909	AZT, 3TC, Kaletra, Tenofovir	1	36	-
923	Nevirapine, 3TC, Tenofovir	1	36	46
949	3TC, Kaletra, Tenofovir	1	36	-
919	Kaletra, Efavirenz, Tenofovir	0	36	-
926	3TC, Efavirenz, Tenofovir	1	36	-
1010	AZT, 3TC, Kaletra, Abacavir, Tenofovir	1	37	-
1020	ddI , Kaletra, Tenofovir	1	37	-
976	ddI , Efavirenz, Tenofovir	1	37	-
1011	Abacavir, Efavirenz, Tenofovir	1	37	-
994	Kaletra, Abacavir, Tenofovir	1	37	-
1230	3TC, Ritonavir, Abacavir, Atazanavir	1	39	-
1071	Efavirenz, Tenofovir, Emtricitabine	1	39	-
1227	Ritonavir, Efavirenz, Tenofovir, Emtricitabine, Atazanavir	1	40	-
1245	3TC, Ritonavir, Abacavir, Tenofovir, Atazanavir	1	40	-
1303	ddI , Ritonavir, Tenofovir, Atazanavir	1	40	-
1222	Ritonavir, Tenofovir, Emtricitabine, Atazanavir	1	40	-
1128	Nevirapine, Tenofovir, Emtricitabine	1	40	-
1253	Kaletra, Tenofovir, Emtricitabine	1	41	-
1342	Ritonavir, Tenofovir, Emtricitabine, Lexiva	1	42	-
10006		0	6	16
10026		0	26	46
10027		0	27	45
10028		0	28	45
10030		1	30	43
10031		0	31	-
10035		0	35	49
10037		1	37	-
10038		0	38	-
10040		0	40	-
10041		1	41	-
10042		1	42	-
10043		1	43	-
10046		1	46	-

10048	1	48	-
10049	1	49	-

B Estimation Appendix

B.1 k-means Clustering Algorithm

We implement the following version of the k-means algorithm. At every period t:

1. We select the products that have not yet being applied the *exit switching* rule. In other words, we select products that are still available for people to swith into at period t. Denote this set of products available for clustering at t, $\mathcal{A}_t$.

2. We re-scale the characteristics of all products available for clustering at t. In order to do this we compute
$$\tilde{\theta}^r = \frac{\theta^r}{\max_{\delta \in \mathcal{A}_t} |\delta^r|}, \text{ for } r = h, x$$
Therefore, by construction $\tilde{\theta}^r \in [-1, 1]$.

3. We choose the first k centroids using k initial $\tilde{\theta}$'s in $\mathcal{A}_t$ randomly selected.

4. We allocate all remaining points in $\mathcal{A}_t$ sequentially. At each step the point selected is the one that is closest to one of the existing clusters. This point is then allocated to the correspondent cluster and the centroid of the cluster is updated. This process is repeated until all points are allocated to a cluster.

5. We undertake a reallocation step in which, taken the centroids as given, all points are allocated to their closest centroid.

6. We calculate the value of (4) for the current allocation.

7. We repeat the process 200 times using different random initial $\tilde{\theta}$'s in $\mathcal{A}_t$. The allocation with the lowest value of (4) is chosen. When simulating clusters in estimation we only repeat the process 50 times to speed up the process.

B.2 GMM Estimation Algorithm

Using the fact that we observe the underlying stochastic process that generates the stochastic process of cluster characteristics we can write the moment condition in equation (31) can be

written as

$$\mathbb{E}\left\{ w\left(z_{it}\right) \otimes \left[\begin{array}{c} \vdots \\ \ln\left(\frac{p_{oit}(z_{it})}{p_{jit}(z_{it})}\right) + \mathbb{E}_{\mathcal{P}}\left[y_{jit}\right] - \mathbb{E}_{\mathcal{P}}\left[y_{oit}\right] \\ + \sum_{s=1}^{T^*} \beta^s P_j^{o(s-1)}\left(z_{it}\right) \\ \mathbb{E}_{\mathcal{P}}\left[D_{it+s}\left(z_{it+s}\right) \sum\limits_{k \in C_{t+s}} p_{kit}\left(z_{it+s}\right)\left[y_{kit+s}\left(z_{it+s}\right) + \psi_{kit+s}\left(z_{it+s}\right)\right] \middle| \cdot, j, \mathcal{P}_t \right] \\ - \sum_{s=1}^{T^*} \beta^s P_o^{o(s-1)}\left(z_{it}\right) \\ \mathbb{E}_{\mathcal{P}}\left[D_{it+s}\left(z_{it+s}\right) \sum\limits_{k \in C_{t+s}} p_{kit}\left(z_{it+s}\right)\left[y_{kit+s}\left(z_{it+s}\right) + \psi_{kit+s}\left(z_{it+s}\right)\right] \middle| \cdot, o, \mathcal{P}_t \right] \\ \vdots \end{array} \right] \right\} = 0 \tag{40}$$

Equation (40) is crucial for our simulation estimation method explained below. The key fact is that we observe the characteristics of the underlying process of product evolution and we are then able to use it to generate the stochastic evolution of clusters. We undertake simulation in order to obtain the value of

$$\begin{array}{c} \sum_{s=1}^{T^*} \beta^s P_j^{o(s-1)}\left(z_{it}\right) \times \\ \mathbb{E}_z\left[D_{it+s}\left(z_{it+s}\right) \sum\limits_{k \in C_{t+s}} p_{kit+s}\left(z_{it+s}\right)\left[y_{kit+s}\left(z_{it+s}\right) + \psi_{kit+s}\left(z_{it+s}\right)\right] \middle| z_{it}, d_{jit} = 1, S_{it}^{(s-1)} = 1, d_i^o, \mathcal{P}_t \right] \end{array} \tag{41}$$

for each individual i and choice j at every period t. Let NS denote the number of simulated technology paths for each individual at every period and let the superscript ns indicate that a quantity has being simulated. For individual i and decision j at period t we write the simulated counterpart of equation (41) as

$$\begin{aligned} & \frac{1}{NS} \sum_{ns} \sum_{s=1}^{T^*} \beta^s P_j^{o(s-1),ns}\left(z_{it}\right) D_{it+s}\left(z_{it+s}^{ns,j}\right) \sum_{k \in C_{t+s}^{ns,j}} d_{kit+s}^{ns,j}\left(z_{it+s}^{ns,j}\right)\left[y_{kit+s}\left(z_{it+s}^{ns,j}\right) + \psi_{kit+s}\left(z_{it+s}^{ns,j}\right)\right] \\ = & \frac{1}{NS} \sum_{ns} \sum_{s=1}^{T^*} \beta^s \left(\prod_{\tau=1}^{s} D_{it+\tau}\left(z_{it+\tau}^{ns,j}\right) \right) \sum_{k \in C_{t+s}^{ns,j}} d_{kit+s}^{ns,j}\left(z_{it+s}^{ns,j}\right)\left[y_{kit+s}\left(z_{it+s}^{ns,j}\right) + \psi_{kit+s}\left(z_{it+s}^{ns,j}\right)\right] \end{aligned} \tag{42}$$

For a given vector of parameters of the utility function, the above simulation must be undertaken NS times for each individual i available at period t, and for all t, and for $J-1$ choices as well as for choice o, which means it must be repeated at least $NS \times T \times N \times J$. Further, notice that within each individual simulation we must simulate N optimal paths, one for every person, in order to obtain the aggregate behavior. In other words, even though we simulate only $NS \times T \times N \times J$ technology paths, we simulate $NS \times T \times N \times J \times N$ individual paths. Given our numbers we will be simulating at most $NS \times 33 \times 1669 \times 6 = NS \times 330,462$

technology paths of length T^* and $NS \times 33 \times 1669 \times 6 \times 1669 = NS \times 551,541,078$ individual paths of length T^*. Relying on Hotz et al. (1994) we could set $NS = 1$ and still obtain consistency. We set $NS = 10$ after trying different values of NS for robustness.

The sample moment conditions will then be

$$\frac{1}{\sum_i \sum_t \delta_{it}} \sum_{i=1}^{N} \sum_{t=1}^{T} \delta_{it} w\left(z_{it}\right) \otimes \left[\begin{array}{c} \vdots \\ \ln\left(\frac{p_{oit}(z_{it})}{p_{jit}(z_{it})}\right) + y_{jit} - y_{oit} \\ + \frac{1}{NS} \sum_{ns} \sum_{s=1}^{T^*} \beta^s \left(\prod_{\tau=1}^{s} D_{it+\tau}\left(z_{it+\tau}^{ns,j}\right) \right) \times \\ \sum_{k \in C_{t+s}^{ns,j}} d_{kit+s}^{ns,j}\left(z_{it+s}^{ns,j}\right) \left[y_{kit+s}\left(z_{it+s}^{ns,j}\right) + \psi_{kit+s}\left(z_{it+s}^{ns,j}\right) \right] \\ - \frac{1}{NS} \sum_{ns} \sum_{s=1}^{T^*} \beta^s \left(\prod_{\tau=1}^{s} D_{it+\tau}\left(z_{it+\tau}^{ns,o}\right) \right) \times \\ \sum_{k \in C_{t+s}^{ns,j}} d_{kit+s}^{ns,o}\left(z_{it+s}^{ns,o}\right) \left[y_{kit+s}\left(z_{it+s}^{ns,o}\right) + \psi_{kit+s}\left(z_{it+s}^{ns,o}\right) \right] \\ \vdots \end{array} \right] = 0 \qquad (43)$$

where δ_{it} is an indicator of availability of individual i at period t. Estimation follows the simulation strategy described below. Simulation will be undertaken in order to obtain

$$\sum_{s=1}^{T^*} \beta^s P_j^{o(s-1)}\left(z_{it}\right) \mathbb{E}_z \left[D_{it+s}\left(z_{it+s}\right) \sum_{k \in C_{t+s}} p_{ikt+s}\left(z_{it+s}\right) \left[y_{ikt+s}\left(z_{it+s}\right) + \psi_{ikt+s}\left(z_{it+s}\right) \right] \middle| z_{it}, d_{jit} = 1, S_{it}^{(s-1)} = 1, d_i^o \right] \qquad (44)$$

and

$$\sum_{s=1}^{T^*} \beta^s P_o^{o(s-1)}\left(z_{it}\right) \mathbb{E}_z \left[D_{it+s}\left(z_{it+s}\right) \sum_{k \in C_{t+s}} p_{ikt+s}\left(z_{it+s}\right) \left[y_{ikt+s}\left(z_{it+s}\right) + \psi_{ikt+s}\left(z_{it+s}\right) \right] \middle| z_{it}, d_{iot} = 1, S_{it}^{(s-1)} = 1, d_i^o \right] \qquad (45)$$

for each individual i at every period t. Let the superscript ns indicate that a quantity has been simulated. Also let subscript j denote the decision made at time t to be compared against the base choice o.

For individual i at period t who chose j, the simulation algorithm to obtain (44) entails the following steps for each simulated path ns (again, we set the number of simulated paths for every data point (i,t) at $NS = 1$):

1. **Number of new products**. If $s = 1$, define $MaxChange_{t+s-1}^{ns} \equiv MaxChange_t$. Using $MaxChange_{t+s-1}^{ns}$ we draw number of new products, New_{t+s}^{ns}, using a negative binomial process. First we draw

$$\mu_{t+s}^* \sim Gamma\left(1/\alpha, \alpha\mu_{t+s}\right)$$

where

$$\mu_{t+s} = \beta_0^N + \beta_1^N MaxChange_{t+s-1}$$

Then we draw

$$New_{t+s}^{ns}\big|_{\mu^*} \sim Poisson\left(\mu_{t+s}^*\right)$$

$\left(\alpha, \beta_0^N, \beta_1^N\right)$ are parameters estimated in a first stage.

2. **Characteristics of new products**. If $New_{t+s}^{ns} > 0$, for each simulated new product we obtain simulated product characteristics. Consistent with our model, new products at $t+s$ are characterized by simulated realizations of the bivariate random vector

$$\omega_{t+s-1} + \nu_{t+s-1} \tag{46}$$

where ω_{t+s-1} is the centroid at $t+s-1$, $\nu_{t+s-1} \sim F_\nu$ and F_ν is our innovations distribution which is estimated non parametrically.

As a by-product of steps 1 and 2 we obtain $MaxChange_{t+s}^{ns}$ using equation (23).

3. **Exit**.

   ↪ *Overall exit rule.* If the ratio of people consuming product k (either by staying or switching) relative to the number of people consuming a market product falls bellow $\tilde{\sigma}_2$ during the last 2 consecutive periods (i.e. $t+s-1, t+s-2$), the product is withdrawn from the market and cannot be consumed at any $\tau \geq t+s$. $\tilde{\sigma}_2$ is chosen as the minimum conditional share observed in the data.

   ↪ *Switching exit rule.* If the product satisfies the overall exit rule or if the ratio of people switching and being assigned product k relative to the number of people switching falls bellow $\tilde{\sigma}_1$ during the last 3 consecutive periods (i.e. $t+s-1, t+s-2, t+s-3$), the product is no longer available for switchers and therefore cannot be used to form clusters at any $\tau \geq t+s$. $\tilde{\sigma}_1$ is chosen as the minimum conditional share observed in the data. These products may still be used by "staying" individuals who consumed the product last period.

   Old products minus exits plus simulated new products yields the simulated set of products in period $t+s$, $\mathcal{P}_{t+s}^{ns}$.

4. **Clusters.** From the simulated set of products $\mathcal{P}_{t+s}^{ns}$, we select those products that can be used for clustering and along with the grouping algorithm we obtain simulated clusters $\mathcal{G}_{t+s}^{ns}$. We then compute characteristics for the simulated clusters, W_{t+s}^{ns}.

5. **Centroid.** If $s = 1$ define $\mathcal{P}^{ns}_{t+s-1} \equiv \mathcal{P}_t$. Using the characteristics of products in $\mathcal{P}^{ns}_{t+s-1}$, unconditional choice probabilities ($\mathbb{E}_i\left[p_{jit+s-1}\left(z_{it}\right)\right]$), within-cluster product weights at $t+s-1$, and $t+s-1$ shares of products conditional on staying, we compute the simulated centroid ω^{ns}_{t+s} using equation (20).

   Steps 1 through 5 provide the aggregate part of the simulated state, $z^{\mathcal{P},ns}_{t+s}$. Denote the future choice set induced by the simulated evolution of products as $\mathcal{C}^{ns}_{t+s}$.

6. **Future state for** i. **(i)** If $s = 1$, define $h^{ns}_{jit+s-1}$as the observed $h_{jit+s-1}$. If $s > 1$, draw $\epsilon^{h,ns}_{it+s-1}$ from the non parametric distribution of ϵ^h; then, using d^{ns}_{it+s-1}, and when necessary, the realization of the within cluster treatment assigned at $t+s-1$, we compute simulated health at the beginning of period $t+s$, $h^{ns}_{jit+s-1}$, using equation (34). If d^{ns}_{it+s-1} involves the trial alternative, trial-product characteristics for computing equation (34) are drawn from the trial distribution at $t+s-1$, $F_{\theta|\omega_{t+s-1}}$; which is equivalent to using equation (46) and the innovations distribution, F_ν. **(ii)** We draw a simulated out-of-pocket payment shock $\epsilon^{o,ns}_{it+s} \sim N\left(0, \sigma^2_o\right)$. **(iii)** We draw a simulated labor state l^{ns}_{it+s} using equation (13). **(iv)** We compute deterministic state variables for i.

7. **Future state for all** $i' \neq i$. **(i)** If $s = 1$, define $h^{ns}_{i't+s-1}$as the observed $h_{i't+s-1}$. If $1 < s < T^*$, draw $\epsilon^{h,ns}_{i't+s-1}$ from the non parametric distribution of ϵ^h. Then, using $d^{ns}_{i't+s-1}$, and when necessary, the realization of the within cluster treatment assigned at $t+s-1$, we compute simulated health at the beginning of period $t+s$, $h^{ns}_{i't+s-1}$, using equation (34). If $d^{ns}_{i't+s-1}$ involves the trial alternative, trial-product characteristics for computing equation (34) are drawn from the trial distribution at $t+s-1$, $F_{\theta|\omega_{t+s-1}}$. We have deliberately written $h^{ns}_{i't+s-1}$ instead of $h^{ns}_{i'jt+s-1}$ as it is explained below. **(ii)** We draw a simulated labor state $l^{ns}_{i't+s}$ using equation (13). **(iii)** We compute deterministic state variables for i'.

   Steps 6 and 7 provide the relevant pieces of the individual-specific part of the simulated state, $z^{o,ns}_{jit+s}$ for i and $z^{o,ns}_{i't+s}$ for all $i' \neq i$.

8. **Probability of Survival up to** $t+s-1$**.** If $s = 1$, by definition, $P^{o(s-1)}_j\left(z_{it}\right) = 1$ for all i available at t. If $s > 1$, using $z^{o,ns}_{jit+s-1}$, and $P^{o(s-2),ns}_j\left(z_{it}\right)$ we obtain $P^{o(s-1),ns}_j\left(z_{it}\right)$ using

$$
\begin{aligned}
P_j^{o(s-1),ns}\left(z_{it}\right) &= \prod_{\tau=1}^{s-1} D_{it+\tau}\left(z_{it+\tau}^{ns}\right) \\
&= D_{it+s-1}\left(z_{jit+s-1}^{ns}\right) P_j^{o(s-2),ns}\left(z_{it}\right) \qquad (47)
\end{aligned}
$$

9. **Conditional choice probabilities and simulated choice for** i. Using $z_{t+s}^{\mathcal{P},ns}$, $z_{jit+s}^{o,ns}$, and equations (36), (38), and (37), we compute simulated $t+s$ ccps, $p_{ikt+s}^{ns}\left(z_{jit+s}^{ns}\right)$, for every alternative $k \in C_{t+s}^{ns}$. Then, using the simulated ccps we draw a decision $d_{it+s}^{ns}\left(z_{jit+s}^{ns}\right)$ for i.

10. **Conditional choice probabilities and simulated choice for all** $i' \neq i$**.** Using $z_{t+s}^{\mathcal{P},ns}$, $z_{i't+s}^{o,ns}$ for all $i' \neq i$, and equations (36), (38), and (37), we compute simulated $t+s$ ccps, $p_{i'kt+s}^{ns}\left(z_{i't+s}^{ns}\right)$, for every alternative $k \in C_{t+s}^{ns}$. Then, using the simulated ccps we draw a decision $d_{i't+s}^{ns}\left(z_{i't+s}^{ns}\right)$ for all $i' \neq i$.

11. **Static payoff for** i**.** **(i)** We compute $\bar{m}_{it+s}^{s} = X_{it+s}^{m,ns}\theta^{m} + \nu_i^m$ using equation (11). Even though individuals know their idiosyncratic shocks in the income equation, ϵ_{it}^{m}, we do not need to simulate these as they are iid and have mean zero an enter linearly in the flow utility, which will result in them averaging out to zero in the moment condition. **(ii)** Using the simulated choice $d_{it+s}^{ns}\left(z_{jit+s}^{ns}\right)$ we compute expected simulated out-of-pocket payments using

$$
o_{it+s}\left(d_{it+s}^{ns}\right) = \begin{cases} o_{it+s}^{*,ns} & \text{if } o_{it+s}^{*,ns} > 0 \\ 0 & \text{if } o_{it+s}^{*,ns} \leq 0 \end{cases}
$$

where

$$
o_{it+s}^{*,ns}\left(d_{it+s}^{ns}\right) = X_{it+s}^{o,ns}\left(d_{it+s}^{ns}\right)\theta^{o} + \epsilon_{it+s}^{o,ns}
$$

and $X_{it+s}^{o,ns}\left(d_{it+s}^{ns}\right)$ are given in equation (12). Hence

$$
\mathbb{E}\left[o_{it+s}\left(d_{it+s}^{ns}\right)\middle|\, d_{it+s}^{ns}\right] = \Phi\left(X_{it+s}^{o,ns}\left(d_{it+s}^{ns}\right)\theta^{o}/\sigma^{o}\right) X_{it+s}^{o,ns}\left(d_{it+s}^{ns}\right)\theta^{o} + \sigma^{o}\phi\left(X_{it+s}^{o,ns}\left(d_{it+s}^{ns}\right)\theta^{o}/\sigma^{o}\right)
$$

(iii) We compute the expected probability of no-ailments as

$$
\mathbb{E}\left[x_{it+s}\middle|\, d_{it+s}^{ns}\right]
$$

using equation (15) and the relevant distribution: cluster, trial, or degenerate. Notice that here we exploit again the fact that we observe the underlying stochastic process.

Whenever the choice is a cluster, we use the within cluster weights. **(iv)** Using above components and i's simulated decision we compute flow payoffs $y^{ns}_{it+s}\left(z^{ns}_{it+s}, d^{ns}_{it+s}\right)$ using equation (9). **(v)** We compute the probability of survival from $t+s-1$ into $t+s$, $D_{it+s}\left(z^{ns}_{it+s}\right)$, using equation (17) and the term $\psi_{it+s}\left(z^{ns}_{it+s}, d^{ns}_{it+s}\right)$ using equation (28).

12. **Repeat all steps above until $s = T^*$.**

In order to obtain all other simulated counterparts of (44) for individual i at period t we do not repeat all the steps above. Instead, we use the same simulated aggregate evolution of the market and repeat only those steps involving individual i's path conditional on choice $j' \neq j$ at t; this is the reason why we deliberately write $h^{ns}_{i't+s-1}$ instead of $h^{ns}_{i'jt+s-1}$ for all $i \neq i'$, as their simulated individual paths do not depend on i's decision at period t. We abstain from generating a path of product innovation following counterfactual choice k by individual i as the impact of his decision at period t on the overall aggregate evolution of the market is negligible.

When simulating the path following counterfactual choice j' we need counterfactual health when $s = 1$, $h^{ns}_{ij't+s-1}$; for this we need to compute the realized residuals of the health equation at t

$$\hat{\epsilon}^h_{it} = h_{it} - \sum_{m=0}^{5} \alpha^h_m h^m_{it-1} - \sum_r \theta^h_r \delta_{it-1r}$$

Then, using the realized residual $\hat{\epsilon}^h_{it}$ and equation (34) we obtain $h^{ns}_{ij't}$. When individual i is in a trial in period t we do not observe the characteristics of the trial ex post; hence, we draw a health shock as well as trial characteristics and compute future simulated health, $h^{ns}_{ij'}$.

Current period payoffs. On the one hand, in order to obtain y_{jit} we need $\mathbb{E}_j\left[x_{jit+s}\right]$. Here, when j corresponds to a cluster alternative, we exploit again the fact that we observe the underlying stochastic process and use the within cluster weights. On the other hand, in order to obtain counterfactual y_{ikt} we need the realized error term of the out-of-pocket payment equation at t given by

$$\hat{\epsilon}^o_{it} = o^*_{jit} - X^o_{jit}\theta^o$$

However, we only observe o^*_{jit} if $o^*_{jit} > 0$. Hence, if $o^*_{jit} \leq 0$, we need to draw a simulated error $\epsilon^{o,ns}_{it}$ from a truncated normal conditional on

$$\epsilon^{o,ns}_{it} \leq -X^o_{jit}\theta^o$$

The sample simulated counterpart of (44) is

$$\frac{1}{NS}\sum_{ns}\sum_{s=1}^{T^*}\beta^s P_j^{o(s-1),ns}\left(z_{it}\right)D_{it+s}\left(z_{it+s}^{ns,j}\right)\sum_{k\in C_{t+s}^{ns,j}}d_{ikt+s}^{ns,j}\left(z_{it+s}^{ns,j}\right)\left[y_{ikt+s}\left(z_{it+s}^{ns,j}\right)+\psi_{ikt+s}\left(z_{it+s}^{ns,j}\right)\right] \quad (48)$$

One potential issue with our simulation algorithm is that in reality individuals die and others become potential consumers. This two phenomena are likely to affect the aggregate joint distribution of individual characteristics and therefore the ccps and the evolution of the market. In order to control for death when computing i's continuation value we could simulate death conditional on optimal behavior for all $i' \neq i$, i.e. some people will leave the sample in the simulated paths. However, we would also need to create people to be introduced into the market. We decide to simulate neither people into the absorbing state nor the stream of people into the sample. Instead, we condition on the aggregate distribution of characteristics at any period t in order to simulate ahead and on optimal future behavior.[22]

Also, a related issue is that our sample is refreshed at least once as new subjects are surveyed. Figures not shown here present no special effect of this refreshing in terms of the aggregate ccps suggesting that the aggregate distribution of characteristics of the new surveyed people matches that of the surveyed individuals at the time.

B.3 Estimator

We use a GMM estimator to obtain our structural parameters. Define B as the $K-$dimensional vector of parameters.Following Hotz et al. (1994) we want to obtain the parameter vector that solves

$$\left((NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\delta_{it}w\left(z_{it}\right)\otimes\bar{v}_{it}(z_{it},B)\right)'W_n\left((NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\delta_{it}w\left(z_{it}\right)\otimes\bar{v}_{it}(z_{it},B)\right) \quad (49)$$

[22] We may need to elaborate here.

where

$$\bar{v}_{it}(z_{it},B)=\left[\begin{array}{c}\vdots\\ \ln\left(\frac{p_{oit}(z_{it})}{p_{jit}(z_{it})}\right)+y_{jit}-y_{oit}\\ +\frac{1}{NS}\sum_{ns}\sum_{s=1}^{T^*}\beta^s\left(\prod_{\tau=1}^{s}D_{it+\tau}\left(z_{it+\tau}^{ns,j}\right)\right)\times\\ \sum_{k\in C_{t+s}^{ns,j}}d_{kit+s}^{ns,j}\left(z_{it+s}^{ns,j}\right)\left[y_{kit+s}\left(z_{it+s}^{ns,j}\right)+\psi_{kit+s}\left(z_{it+s}^{ns,j}\right)\right]\\ -\frac{1}{NS}\sum_{ns}\sum_{s=1}^{T^*}\beta^s\left(\prod_{\tau=1}^{s}D_{it+\tau}\left(z_{it+\tau}^{ns,o}\right)\right)\times\\ \sum_{k\in C_{t+s}^{ns,j}}d_{kit+s}^{ns,o}\left(z_{it+s}^{ns,o}\right)\left[y_{kit+s}\left(z_{it+s}^{ns,o}\right)+\psi_{kit+s}\left(z_{it+s}^{ns,o}\right)\right]\\ \vdots\end{array}\right]$$

and W_n is a square weighting matrix. Using the linear structure of our utility function we collect and factor terms in order to write the jth component of the vector $\bar{v}_{it}(z_{it},B)$ as the linear form

$$\tilde{y}_{jit}-\tilde{x}_{jit}'B$$

Define Y as the $[(J-1)NT\times 1]$ −dimensional vector that stacks all $\tilde{y}_{jit}$, X the matrix of dimensions $[(J-1)NT\times K]$ that stacks all $\tilde{x}_{jit}$. Define Z as the $[NT\times R]$ −dimensional matrix the columns of which contain the R instruments orthogonal to the difference in alternative representations—which renders W_n as a $(J-1)R-$dimensional square matrix. Finally, let $\mathbf{I}_{[J-1]}$ be a $(J-1)-$dimensional identity matrix

$$Y=\left[\begin{array}{c}\tilde{y}_{1,1,1}\\ \tilde{y}_{1,1,2}\\ \vdots\\ \tilde{y}_{1,N,T-1}\\ \tilde{y}_{1,N,T}\\ \vdots\\ \tilde{y}_{J-1,1,1}\\ \tilde{y}_{J-1,1,2}\\ \vdots\\ \tilde{y}_{J-1,N,T-1}\\ \tilde{y}_{J-1,N,T}\end{array}\right],\quad X=\left[\begin{array}{ccc}\tilde{x}_{1,1,1,1}&\cdots&\tilde{x}_{1,1,1,K}\\ \tilde{x}_{1,1,2,1}&\cdots&\tilde{x}_{1,1,2,K}\\ \vdots&&\vdots\\ \tilde{x}_{1,N,T-1,1}&\cdots&\tilde{x}_{1,N,T-1,K}\\ \tilde{x}_{1,N,T,1}&\cdots&\tilde{x}_{1,N,T,K}\\ \vdots&&\vdots\\ \tilde{x}_{J-1,1,1,1}&\cdots&\tilde{x}_{J-1,1,1,K}\\ \tilde{x}_{J-1,1,2,1}&\cdots&\tilde{x}_{J-1,1,2,K}\\ \vdots&&\vdots\\ \tilde{x}_{J-1,N,T-1,1}&\cdots&\tilde{x}_{J-1,N,T-1,K}\\ \tilde{x}_{J-1,N,T,1}&\cdots&\tilde{x}_{J-1,N,T,K}\end{array}\right],\quad Z=\left[\begin{array}{ccc}w\left(z_{11}\right)_1&\cdots&w\left(z_{11}\right)_R\\ w\left(z_{12}\right)_1&\cdots&w\left(z_{12}\right)_R\\ \vdots&&\vdots\\ w\left(z_{NT}\right)_1&\cdots&w\left(z_{NT}\right)_R\end{array}\right]$$

And define

$$\tilde{Z}=\mathbf{I}_{[J-1]}\otimes Z$$

Then we can write the objective function in (49) as

$$\left((NT)^{-1}\tilde{Z}'\left(Y-XB\right)\right)'W_n\left((NT)^{-1}\tilde{Z}'\left(Y-XB\right)\right)$$

From where we can obtain a close form solution for $\hat{B}$ as the optimal GMM estimator. It entails a first stage estimator given by

$$\hat{B}^{1S} = \left(X'\tilde{Z}\tilde{Z}'X\right)^{-1}\left(X'\tilde{Z}\tilde{Z}'Y\right)$$

and a second stage estimator given by

$$\hat{B}^{2S} = \left(X'\tilde{Z}\hat{S}^{-1}\tilde{Z}'X\right)^{-1}\left(X'\tilde{Z}\hat{S}^{-1}\tilde{Z}'Y\right)$$

where

$$\hat{S} = \frac{1}{N^*}\tilde{Z}'D\tilde{Z}$$

and D is the $N(J-1) \times N(J-1)$ diagonal matrix with elements $\hat{u}_{jit}^2 = \left(y_{jit} - x'_{jit}\hat{B}^{1S}\right)^2$ in its diagonal. The variance-covariance matrix of the second stage estimator is

$$\hat{V}^{2S} = N^*\left(X'\tilde{Z}\hat{S}^{-1}\tilde{Z}'X\right)^{-1}$$

and

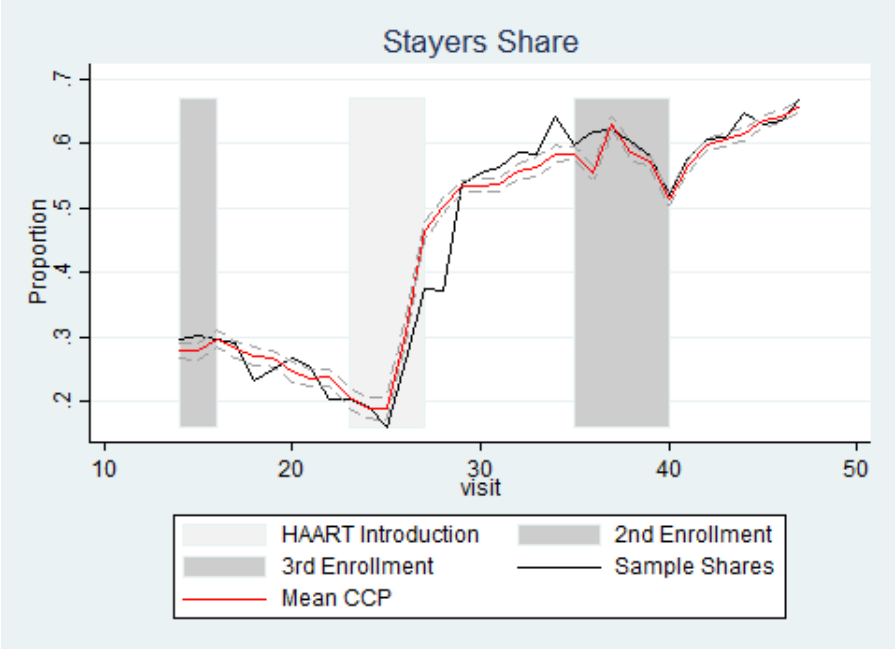
$$N^* = \sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{J-1} 1\{\text{Decision } j \text{ available for } i \text{ at } t\}$$

which accounts for the fact that some individuals cannot stay in their lagged treatments at some periods (for instance, if lagged decision was no treatment or trial treatment).

B.4 CCP Estimation Fit

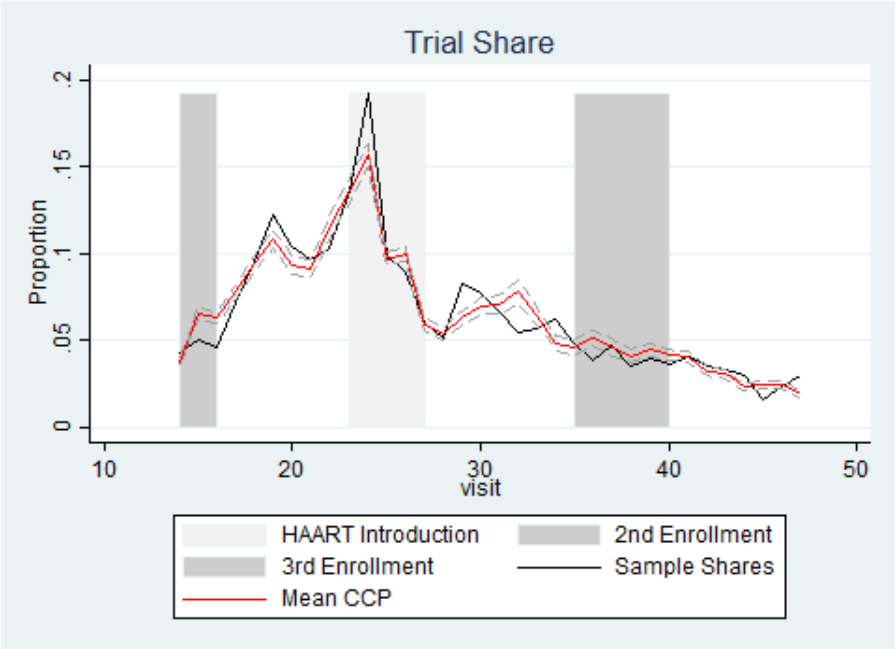
Figures S1, S2, and S3 display the mean predicted conditional choice probability using equations (36), (38), (37) and (39) over time against the correspondent share of the population who chose the alternative. Our ccps map the choices in the data relatively well. In fact, we further explore the fit of our ccp estimates comparing the relatives shares that clusters received in reality against our the predictions from our estimated ccps. We do this by ranking the three clusters at every period by the share they received and comparing this ranking with the ranking obtained from our estimated ccps. A cross tabulation of these rankings—not shown here—suggests that the predicted ranks match the real ranks in more than 79 percent of the periods. In fact, the lowest-ranking cluster matches the predicted lowest-ranking cluster 88 percent of the times.

Appendix Figure S1: CCPs Goodness of fit: Stayers



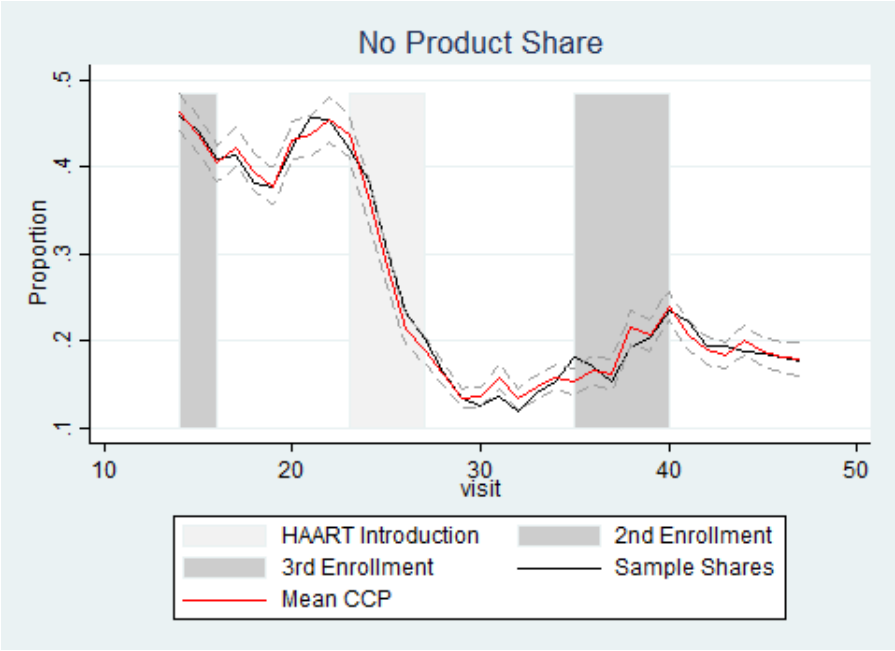
Notes: Figure shows the average estimated conditional choice probability against the share of people choosing the alternative. Dashed lines represent 95 percent confidence intervals around the predicted conditional choice probabilities. Figure highlights three periods that are of special relevance in the period we study. It depicts periods during which enrollment into the sample was undertaken and more importantly, it displays the period in which products belonging to the HAART family were introduced.

Appendix Figure S2: CCPs Goodness of fit: Trial



Notes: Figure shows the average estimated conditional choice probability against the share of people choosing the alternative. Dashed lines represent 95 percent confidence intervals around the predicted conditional choice probabilities. Figure highlights three periods that are of special relevance in the period we study. It depicts periods during which enrollment into the sample was undertaken and more importantly, it displays the period in which products belonging to the HAART family were introduced.

Appendix Figure S3: CCPs Goodness of fit: No Product



Notes: Figure shows the average estimated conditional choice probability against the share of people choosing the alternative. Dashed lines represent 95 percent confidence intervals around the predicted conditional choice probabilities. Figure highlights three periods that are of special relevance in the period we study. It depicts periods during which enrollment into the sample was undertaken and more importantly, it displays the period in which products belonging to the HAART family were introduced.

C Proofs

C.1 Proof of Proposition 1

$$
\begin{aligned}
v_{jit}(z_{it}) &= y_{jit} + \beta \mathbb{E}\left[V\left(z_{it+1}, \varepsilon_{it+1}\right) | z_{it}, j\right] \\
&= y_{jit} + \beta \mathbb{E}\left[D_{it+1}\left(z_{it+1}\right) \mathbb{E}_{\varepsilon}\left[\sum_{k \in C_{t+1}} d^{o}_{ikt+1}\left(z_{it+s}\right)\left[y_{ikt+s}\left(z_{it+s}\right) + \varepsilon_{ikt+1}\right]\right] | z_{it}, j\right] \\
&\quad + \beta^{2} \mathbb{E}\left[D_{it+1}\left(z_{it+1}\right) V\left(z_{it+2}, \varepsilon_{it+2}\right) | z_{it}, j, d^{o}_{i}\right] \\
&= y_{jit} + \beta \mathbb{E}\left[D_{it+1}\left(z_{it+1}\right) \mathbb{E}_{\varepsilon}\left[\sum_{k \in C_{t+1}} \mathbb{E}_{\varepsilon}\left[d^{o}_{ikt+1}\left(z_{it+s}\right)\left[y_{ikt+s}\left(z_{it+s}\right) + \varepsilon_{ikt+1}\right] \middle| d^{o}_{ikt+1}\left(z_{it+s}\right) = 1\right]\right] | z_{it}, j\right] \\
&\quad + \beta^{2} \mathbb{E}\left[D_{it+1}\left(z_{it+1}\right) V\left(z_{it+2}, \varepsilon_{it+2}\right) | z_{it}, j, d^{o}_{i}\right] \\
&= y_{jit} + \beta \mathbb{E}\left[D_{it+1}\left(z_{it+1}\right) \mathbb{E}_{\varepsilon}\left[\sum_{k \in C_{t+1}} d^{o}_{ikt+1}\left(z_{it+s}\right)\left[y_{ikt+s}\left(z_{it+s}\right) + \mathbb{E}_{\varepsilon}\left[\varepsilon_{ikt+1} \middle| d^{o}_{ikt+1}\left(z_{it+s}\right) = 1\right]\right]\right] | z_{it}, j\right] \\
&\quad + \beta^{2} \mathbb{E}\left[D_{it+1}\left(z_{it+1}\right) V\left(z_{it+2}, \varepsilon_{it+2}\right) | z_{it}, j, d^{o}_{i}\right] \\
&= y_{jit} + \beta \mathbb{E}\left[D_{it+1}\left(z_{it+1}\right) \mathbb{E}_{\varepsilon}\left[\sum_{k \in C_{t+1}} d^{o}_{ikt+1}\left(z_{it+s}\right)\left[y_{ikt+s}\left(z_{it+s}\right) + \psi_{ikt+s}\left(z_{it+s}\right)\right]\right] | z_{it}, j\right] \\
&\quad + \beta^{2} \mathbb{E}\left[D_{it+1}\left(z_{it+1}\right) V\left(z_{it+2}, \varepsilon_{it+2}\right) | z_{it}, j, d^{o}_{i}\right] \\
&= y_{jit} + \beta \mathbb{E}\left[D_{it+1}\left(z_{it+1}\right) \mathbb{E}_{\varepsilon}\left[\sum_{k \in C_{t+1}} d^{o}_{ikt+1}\left(z_{it+1}\right)\left[y_{ikt+1}\left(z_{it+1}\right) + \psi_{ikt+1}\left(z_{it+1}\right)\right]\right] | z_{it}, j\right] \\
&\quad + \beta^{2} P_{j}^{o(2-1)}\left(z_{it}\right) \mathbb{E}\left[V\left(z_{it+2}, \varepsilon_{it+2}\right) | z_{it}, j, S_{it+2-1} = 1, d^{o}_{i}\right] \\
&= y_{jit} \\
&\quad + \sum_{s=1}^{T^{*}} \beta^{s} P_{j}^{o(s-1)}\left(z_{it}\right) \mathbb{E}_{z}\left[D_{it+s}\left(z_{it+s}\right) \sum_{k \in C_{t+s}} p_{ikt+s}\left(z_{it+s}\right)\left[y_{ikt+s}\left(z_{it+s}\right) + \psi_{ikt+s}\left(z_{it+s}\right)\right] \middle| z_{it}, j, S_{it+s-1} = 1, d^{o}_{i}\right] \\
&\quad + \beta^{T^{*}+1} P_{j}^{o(T^{*})}\left(z_{it}\right) \mathbb{E}_{z}\left[D_{it+T^{*}+1}\left(z_{it+T^{*}+1}\right) V\left(z_{it+T^{*}+1}, \varepsilon_{it+T^{*}+1}\right) | z_{it}, j, S_{it+T^{*}} = 1, d^{o}_{i}\right]
\end{aligned}
$$

That

$$
\psi_{kit}\left(z_{it}\right) = \gamma - \ln\left(p_{kit}\left(z_{it}\right)\right) \tag{50}
$$

follows from the joint distribution of the taste shifter ε_{it}, which is Extreme Value Type-I. *Q.E.D.*

D Additional Tables and Figures

Appendix Table S5: Health Characteristics of Treatments

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	cd4	cd4	cd4	cd4	cd4	cd4
$CD4_{t-1}$	0.834***	1.064***	1.009***	1.021***	1.152***	1.136***
	(0.006)	(0.013)	(0.014)	(0.026)	(0.032)	(0.051)
$CD4^2_{t-1}/10^3$		-0.174***	-0.097***	-0.121**	-0.519***	-0.456**
		(0.012)	(0.018)	(0.059)	(0.098)	(0.212)
$CD4^3_{t-1}/10^7$			-0.274***	-0.112	4.375***	3.363
			(0.060)	(0.430)	(1.123)	(3.518)
$CD4^4_{t-1}/10^{10}$				-0.031	-2.016***	-1.288
				(0.080)	(0.502)	(2.598)
$CD4^5_{t-1}/10^{14}$					2.803***	0.482
					(0.718)	(8.325)
$CD4^6_{t-1}/10^{18}$						2.623
						(9.398)
DT3	-21.583***	-11.174***	-11.950***	-11.909***	-12.004***	-11.983***
	(2.840)	(2.707)	(2.700)	(2.699)	(2.697)	(2.697)
DT9	-20.319*	-19.561*	-19.089	-19.165	-19.655*	-19.623*
	(11.743)	(11.879)	(11.865)	(11.865)	(11.861)	(11.860)
DT13	-72.170***	-53.988***	-55.512***	-55.437***	-55.796***	-55.726***
	(11.798)	(12.504)	(12.348)	(12.364)	(12.455)	(12.455)
DT14	-15.506**	-4.315	-5.164	-5.115	-5.155	-5.140
	(6.241)	(6.105)	(6.100)	(6.098)	(6.094)	(6.094)
DT33	-21.629**	0.985	-2.034	-1.689	-0.017	-0.112
	(9.038)	(9.161)	(9.112)	(9.123)	(9.108)	(9.113)
DT34	-21.810***	-12.450***	-13.310***	-13.219***	-12.752***	-12.779***
	(4.900)	(4.755)	(4.761)	(4.764)	(4.764)	(4.765)
DT39	-31.246***	-15.492***	-17.057***	-16.924***	-16.615***	-16.607***
	(5.805)	(5.715)	(5.691)	(5.691)	(5.687)	(5.688)
DT46	7.741*	15.348***	14.581***	14.678***	15.263***	15.229***
	(4.679)	(4.566)	(4.577)	(4.581)	(4.574)	(4.573)
DT47	-34.371***	-15.583**	-17.510**	-17.327**	-16.474**	-16.521**
	(7.318)	(7.033)	(7.027)	(7.028)	(7.040)	(7.037)
DT51	-22.630***	-3.664	-6.022	-5.740	-4.159	-4.252
	(7.962)	(7.693)	(7.690)	(7.693)	(7.669)	(7.670)
DT63	-16.743	2.384	-0.183	0.162	2.415	2.275
	(14.637)	(13.594)	(13.649)	(13.659)	(13.746)	(13.735)
DT64	-37.988***	-17.449*	-19.583**	-19.387**	-18.630**	-18.656**
	(8.900)	(9.076)	(8.991)	(8.996)	(9.035)	(9.032)
DT65	-27.913***	-12.409*	-13.823**	-13.704*	-13.186*	-13.220*
	(7.203)	(7.007)	(7.006)	(7.006)	(6.993)	(6.994)
DT67	-50.755***	-31.179**	-33.087**	-32.938**	-32.700**	-32.673**
	(15.300)	(14.998)	(14.986)	(14.990)	(15.052)	(15.044)
DT69	-26.741*	-11.827	-13.331	-13.215	-13.351	-13.275
	(14.379)	(13.908)	(13.950)	(13.945)	(13.973)	(13.973)
DT85	34.619***	40.457***	39.721***	39.790***	39.776***	39.792***
	(6.424)	(6.319)	(6.311)	(6.308)	(6.299)	(6.299)
DT117	33.323***	42.736***	41.910***	41.991***	42.267***	42.277***
	(12.098)	(11.819)	(11.837)	(11.834)	(11.819)	(11.818)
DT124	33.711***	33.910***	33.804***	33.864***	34.398***	34.364***

	(6.284)	(6.229)	(6.234)	(6.235)	(6.227)	(6.228)
DT146	27.752*	34.323**	33.694**	33.761**	33.792**	33.831**
	(14.796)	(14.611)	(14.644)	(14.639)	(14.625)	(14.625)
DT157	34.353***	37.455***	37.258***	37.282***	37.173***	37.208***
	(6.945)	(6.861)	(6.862)	(6.862)	(6.856)	(6.857)
DT161	33.496**	38.559***	38.364***	38.340***	38.283***	38.259***
	(13.411)	(13.215)	(13.201)	(13.205)	(13.193)	(13.198)
DT164	55.089**	64.825**	64.314**	64.302**	63.734**	63.798**
	(27.560)	(27.365)	(27.365)	(27.370)	(27.414)	(27.419)
DT165	60.168***	64.722***	65.220***	65.337***	65.041***	65.045***
	(7.077)	(6.246)	(6.215)	(6.225)	(6.220)	(6.222)
DT169	33.129**	34.545**	34.182**	34.289**	35.032**	35.012**
	(16.388)	(16.316)	(16.316)	(16.317)	(16.339)	(16.334)
DT171	73.104***	78.825***	78.453***	78.478***	78.559***	78.548***
	(17.682)	(18.040)	(18.017)	(18.010)	(17.950)	(17.943)
DT175	44.728***	52.470***	52.730***	52.619***	53.128***	53.153***
	(8.770)	(8.233)	(8.202)	(8.198)	(8.176)	(8.172)
DT185	50.899***	58.842***	57.833***	57.922***	57.776***	57.825***
	(12.661)	(12.659)	(12.642)	(12.638)	(12.608)	(12.608)
DT202	32.648**	32.522**	33.226**	33.107**	32.286**	32.338**
	(14.544)	(14.584)	(14.576)	(14.576)	(14.573)	(14.574)
DT214	33.330***	33.541***	34.154***	34.057***	33.510***	33.535***
	(12.245)	(12.162)	(12.166)	(12.164)	(12.163)	(12.161)
DT236	48.886***	46.186***	46.239***	46.267***	46.275***	46.281***
	(7.172)	(7.130)	(7.121)	(7.123)	(7.123)	(7.124)
DT242	47.980***	46.484***	46.110***	46.240***	46.846***	46.863***
	(9.266)	(9.144)	(9.154)	(9.160)	(9.161)	(9.162)
DT254	42.775***	42.587***	42.568***	42.601***	42.631***	42.656***
	(13.502)	(13.436)	(13.444)	(13.443)	(13.446)	(13.445)
DT268	47.316***	52.474***	51.249***	51.353***	50.776***	50.855***
	(10.457)	(10.436)	(10.415)	(10.407)	(10.417)	(10.415)
DT292	42.030***	48.796***	48.057***	48.109***	48.018***	48.069***
	(10.638)	(10.242)	(10.225)	(10.230)	(10.212)	(10.215)
DT311	39.770*	50.740**	49.168**	49.215**	47.816**	47.922**
	(23.206)	(22.720)	(22.698)	(22.709)	(22.774)	(22.780)
DT349	50.575***	42.467**	43.413**	43.408**	44.240***	44.156***
	(16.958)	(17.083)	(17.108)	(17.098)	(17.046)	(17.046)
DT377	56.809***	57.778***	57.727***	57.716***	57.227***	57.259***
	(19.614)	(19.566)	(19.530)	(19.538)	(19.552)	(19.555)
DT532	49.321***	47.631***	47.537***	47.614***	47.978***	47.990***
	(10.952)	(10.886)	(10.899)	(10.899)	(10.893)	(10.892)
DT548	45.842***	43.345***	43.281***	43.348***	43.526***	43.551***
	(5.368)	(5.331)	(5.329)	(5.330)	(5.327)	(5.329)
DT573	40.314***	39.595***	39.432***	39.485***	39.379***	39.426***
	(10.981)	(10.960)	(10.901)	(10.909)	(10.919)	(10.922)
DT581	19.612	18.387	18.413	18.417	17.866	17.935
	(14.499)	(14.316)	(14.341)	(14.341)	(14.376)	(14.378)
DT615	44.239***	40.856***	41.087***	41.122***	41.280***	41.301***
	(11.769)	(11.592)	(11.603)	(11.604)	(11.622)	(11.622)
DT644	54.543***	53.883***	53.615***	53.650***	53.341***	53.368***
	(8.639)	(8.512)	(8.504)	(8.505)	(8.516)	(8.515)
DT701	52.853***	55.916***	54.878***	54.997***	54.824***	54.870***
	(11.144)	(10.997)	(10.979)	(10.975)	(10.999)	(11.003)
DT720	60.713***	79.231***	78.995***	78.688***	78.914***	78.726***
	(13.046)	(14.550)	(14.500)	(14.464)	(14.412)	(14.405)

DT782	28.143*	35.924**	35.177**	35.267**	35.611**	35.633**
	(15.627)	(14.945)	(14.959)	(14.967)	(15.077)	(15.068)
DT869	50.005***	50.037***	49.904***	49.946***	49.838***	49.885***
	(13.110)	(12.947)	(12.940)	(12.946)	(12.997)	(12.999)
DT909	27.964**	33.525***	33.367***	33.310***	32.227***	32.320***
	(10.898)	(10.827)	(10.764)	(10.781)	(10.842)	(10.846)
DT919	47.628***	48.846***	48.522***	48.536***	47.617***	47.722***
	(13.308)	(13.619)	(13.486)	(13.499)	(13.453)	(13.466)
DT921	15.458	18.929	19.116	19.058	19.273	19.286
	(16.086)	(14.468)	(14.426)	(14.448)	(14.327)	(14.341)
DT923	30.619*	26.759	26.896	26.960	27.246	27.283
	(17.077)	(16.964)	(16.951)	(16.954)	(16.971)	(16.973)
DT926	47.776***	47.965***	47.952***	47.971***	47.790***	47.835***
	(10.277)	(9.953)	(9.995)	(9.993)	(10.024)	(10.021)
DT940	54.575***	50.873***	50.705***	50.821***	51.570***	51.537***
	(15.098)	(15.196)	(15.198)	(15.195)	(15.164)	(15.163)
DT949	53.752***	50.921***	51.203***	51.236***	51.672***	51.661***
	(11.818)	(11.668)	(11.706)	(11.701)	(11.705)	(11.701)
DT963	37.095***	30.953**	31.641**	31.628**	31.845**	31.829**
	(12.997)	(13.013)	(13.021)	(13.019)	(13.014)	(13.015)
DT976	-7.655	2.510	2.487	2.349	2.381	2.436
	(20.271)	(17.213)	(16.922)	(16.980)	(16.471)	(16.507)
DT987	4.115	10.199	9.583	9.654	9.855	9.870
	(14.474)	(14.526)	(14.496)	(14.498)	(14.503)	(14.500)
DT994	10.625	15.088	15.225	15.171	14.891	14.893
	(17.109)	(17.416)	(17.404)	(17.396)	(17.282)	(17.291)
DT1010	12.409	20.292**	19.891**	19.906**	19.980**	19.966**
	(9.769)	(9.701)	(9.687)	(9.686)	(9.676)	(9.676)
DT1011	43.817*	39.067*	38.498*	38.691*	39.457*	39.484*
	(22.599)	(22.024)	(22.150)	(22.142)	(22.163)	(22.157)
DT1020	20.973	17.603	18.253	18.220	18.396	18.375
	(13.204)	(13.121)	(13.141)	(13.135)	(13.111)	(13.111)
DT1071	58.721***	53.880***	54.012***	54.094***	54.798***	54.796***
	(4.578)	(4.475)	(4.469)	(4.478)	(4.453)	(4.453)
DT1128	40.782***	37.265***	37.397***	37.415***	37.227***	37.246***
	(9.919)	(9.662)	(9.682)	(9.683)	(9.722)	(9.720)
DT1222	53.239***	52.943***	52.719***	52.793***	53.028***	53.050***
	(5.409)	(5.301)	(5.306)	(5.307)	(5.309)	(5.310)
DT1227	83.132***	85.565***	84.437***	84.556***	83.823***	83.917***
	(20.468)	(21.138)	(21.037)	(21.031)	(20.842)	(20.859)
DT1230	25.312**	27.037**	27.504**	27.416**	26.850**	26.886**
	(12.040)	(12.112)	(12.093)	(12.093)	(12.079)	(12.081)
DT1245	37.822***	38.379***	38.357***	38.374***	38.313***	38.336***
	(13.418)	(13.334)	(13.326)	(13.328)	(13.347)	(13.347)
DT1253	46.066***	46.654***	46.213***	46.319***	46.723***	46.735***
	(7.946)	(7.735)	(7.745)	(7.747)	(7.767)	(7.765)
DT1303	51.942***	46.893***	47.631***	47.602***	47.800***	47.786***
	(16.434)	(16.225)	(16.276)	(16.269)	(16.284)	(16.279)
DT1342	32.840**	29.637**	30.481**	30.395**	30.226**	30.204**
	(14.069)	(14.124)	(14.137)	(14.132)	(14.115)	(14.116)
DT10006	-29.515	-21.709	-22.258	-22.209	-21.950	-21.961
	(19.930)	(20.278)	(20.256)	(20.251)	(20.191)	(20.194)
DT10026	64.425***	66.829***	65.194***	65.444***	65.353***	65.496***
	(24.136)	(22.649)	(22.765)	(22.778)	(23.021)	(23.023)
DT10027	0.210	7.741	7.165	7.179	6.457	6.562

	(14.742)	(14.802)	(14.701)	(14.715)	(14.739)	(14.748)
DT10028	26.450	32.481	31.616	31.629	30.293	30.414
	(19.796)	(20.521)	(20.367)	(20.371)	(20.276)	(20.282)
DT10030	20.031	19.261	19.703	19.654	19.278	19.315
	(17.621)	(17.572)	(17.592)	(17.586)	(17.546)	(17.548)
DT10031	34.205**	29.830**	30.781**	30.712**	31.044**	30.969**
	(15.053)	(14.878)	(14.911)	(14.901)	(14.838)	(14.843)
DT10035	44.294***	43.738***	43.648***	43.682***	43.495***	43.550***
	(15.320)	(15.368)	(15.337)	(15.340)	(15.357)	(15.356)
DT10037	28.525***	28.144***	28.115***	28.134***	27.893**	27.931**
	(11.035)	(10.856)	(10.855)	(10.857)	(10.893)	(10.892)
DT10038	48.755***	44.792***	44.735***	44.855***	45.683***	45.679***
	(13.769)	(13.693)	(13.692)	(13.695)	(13.673)	(13.674)
DT10040	26.314***	28.505***	28.742***	28.692***	28.440***	28.431***
	(8.531)	(8.329)	(8.333)	(8.335)	(8.334)	(8.337)
DT10041	43.028***	41.990***	42.154***	42.158***	42.050***	42.074***
	(12.657)	(12.467)	(12.470)	(12.470)	(12.474)	(12.473)
DT10042	30.254**	32.127**	31.743**	31.779**	31.824**	31.811**
	(12.900)	(13.008)	(13.007)	(12.996)	(12.924)	(12.922)
DT10043	28.589**	26.499**	26.919**	26.894**	26.678**	26.715**
	(12.874)	(12.711)	(12.704)	(12.707)	(12.739)	(12.739)
DT10046	36.196***	32.318**	32.455**	32.517**	32.865**	32.884**
	(13.618)	(13.720)	(13.651)	(13.662)	(13.672)	(13.678)
DT10048	33.017**	33.618**	33.783**	33.756**	33.352**	33.394**
	(15.278)	(14.856)	(14.884)	(14.888)	(14.943)	(14.943)
DT10049	49.474***	47.695***	48.034***	48.005***	47.736***	47.760***
	(11.002)	(10.687)	(10.698)	(10.703)	(10.757)	(10.753)
DT10050	38.954**	38.166**	38.394**	38.415**	38.933**	38.887**
	(17.897)	(18.043)	(18.036)	(18.037)	(18.096)	(18.090)
Constant	56.819***	-2.797	6.623**	5.233	-5.874*	-4.898
	(3.161)	(3.411)	(3.018)	(3.432)	(3.313)	(3.681)
Observations	33,258	33,258	33,258	33,258	33,258	33,258
R-squared	0.728	0.736	0.736	0.736	0.736	0.736

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Appendix Table S6: No Ailments Characteristics of Treatments

VARIABLES	(1) NoSfx	(2) NoSfx	(3) NoSfx	(4) NoSfx	(5) NoSfx	(6) NoSfx
$CD4_{t-1}$	0.001***	0.003***	0.005***	0.006***	0.008***	0.009***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)
$CD4^2_{t-1}/10^3$		-0.001***	-0.005***	-0.007***	-0.013***	-0.017***
		(0.000)	(0.001)	(0.001)	(0.002)	(0.003)
$CD4^3_{t-1}/10^7$			0.013***	0.033***	0.109***	0.166***
			(0.003)	(0.004)	(0.019)	(0.040)
$CD4^4_{t-1}/10^{10}$				-0.005***	-0.040***	-0.084***
				(0.001)	(0.010)	(0.028)
$CD4^5_{t-1}/10^{14}$					0.054***	0.203**
					(0.016)	(0.092)
$CD4^6_{t-1}/10^{18}$						-0.186*
						(0.106)
DT3	-0.576***	-0.515***	-0.500***	-0.498***	-0.500***	-0.501***
	(0.040)	(0.041)	(0.041)	(0.041)	(0.041)	(0.041)
DT9	-0.402*	-0.404*	-0.421**	-0.427**	-0.433**	-0.434**
	(0.208)	(0.212)	(0.213)	(0.213)	(0.212)	(0.212)
DT13	-0.738***	-0.627**	-0.594**	-0.593**	-0.600**	-0.603**
	(0.260)	(0.259)	(0.260)	(0.260)	(0.260)	(0.260)
DT14	-0.523***	-0.457***	-0.439***	-0.437***	-0.439***	-0.439***
	(0.099)	(0.100)	(0.100)	(0.101)	(0.101)	(0.101)
DT33	-1.039***	-0.900***	-0.819***	-0.799***	-0.783***	-0.781***
	(0.127)	(0.128)	(0.130)	(0.130)	(0.131)	(0.132)
DT34	-0.622***	-0.569***	-0.549***	-0.544***	-0.539***	-0.538***
	(0.079)	(0.079)	(0.080)	(0.080)	(0.080)	(0.081)
DT39	-0.713***	-0.618***	-0.580***	-0.573***	-0.571***	-0.571***
	(0.105)	(0.106)	(0.107)	(0.108)	(0.108)	(0.108)
DT46	-0.457***	-0.411***	-0.390***	-0.383***	-0.375***	-0.374***
	(0.069)	(0.070)	(0.070)	(0.071)	(0.071)	(0.071)
DT47	-1.016***	-0.907***	-0.866***	-0.858***	-0.851***	-0.850***
	(0.157)	(0.159)	(0.162)	(0.162)	(0.163)	(0.163)
DT51	-0.577***	-0.456***	-0.388**	-0.369**	-0.348**	-0.344**
	(0.161)	(0.164)	(0.166)	(0.167)	(0.168)	(0.168)
DT63	-0.567**	-0.445	-0.367	-0.341	-0.310	-0.304
	(0.269)	(0.274)	(0.280)	(0.282)	(0.284)	(0.285)
DT64	-0.566***	-0.436**	-0.379*	-0.367*	-0.358	-0.357
	(0.210)	(0.213)	(0.218)	(0.220)	(0.222)	(0.222)
DT65	-0.650***	-0.557***	-0.525***	-0.519***	-0.514***	-0.512***
	(0.133)	(0.133)	(0.135)	(0.135)	(0.136)	(0.137)
DT67	-1.582***	-1.473***	-1.439***	-1.436***	-1.440***	-1.442***
	(0.306)	(0.303)	(0.305)	(0.306)	(0.307)	(0.307)
DT69	-0.916***	-0.827***	-0.790***	-0.785***	-0.789***	-0.793***
	(0.261)	(0.261)	(0.262)	(0.263)	(0.263)	(0.264)
DT85	-0.772***	-0.737***	-0.718***	-0.715***	-0.717***	-0.718***
	(0.091)	(0.091)	(0.091)	(0.092)	(0.092)	(0.092)
DT117	-0.609***	-0.555***	-0.535**	-0.530**	-0.527**	-0.528**
	(0.205)	(0.208)	(0.210)	(0.211)	(0.212)	(0.212)
DT124	0.036	0.040	0.050	0.055	0.064	0.065
	(0.093)	(0.094)	(0.094)	(0.094)	(0.094)	(0.094)
DT146	-0.568**	-0.530**	-0.513**	-0.509**	-0.509**	-0.511**
	(0.223)	(0.226)	(0.227)	(0.227)	(0.227)	(0.227)

DT157	-0.132	-0.111	-0.103	-0.102	-0.104	-0.105
	(0.112)	(0.113)	(0.114)	(0.114)	(0.114)	(0.114)
DT161	-0.302*	-0.270	-0.270	-0.272	-0.271	-0.270
	(0.180)	(0.182)	(0.182)	(0.182)	(0.182)	(0.182)
DT164	-0.532***	-0.475**	-0.468**	-0.471**	-0.479**	-0.482**
	(0.184)	(0.185)	(0.186)	(0.187)	(0.187)	(0.187)
DT165	-0.106	-0.084	-0.078	-0.073	-0.075	-0.076
	(0.080)	(0.081)	(0.081)	(0.081)	(0.081)	(0.081)
DT169	-0.604**	-0.600**	-0.585**	-0.577**	-0.567**	-0.566**
	(0.244)	(0.246)	(0.248)	(0.249)	(0.250)	(0.250)
DT171	-0.342*	-0.308	-0.299	-0.297	-0.295	-0.295
	(0.190)	(0.195)	(0.197)	(0.198)	(0.199)	(0.199)
DT175	-0.424***	-0.377***	-0.399***	-0.402***	-0.395***	-0.397***
	(0.099)	(0.100)	(0.100)	(0.100)	(0.100)	(0.100)
DT185	-0.671***	-0.621***	-0.592***	-0.588***	-0.591***	-0.594***
	(0.174)	(0.175)	(0.176)	(0.176)	(0.177)	(0.177)
DT202	0.003	-0.002	-0.027	-0.036	-0.048	-0.050
	(0.226)	(0.227)	(0.228)	(0.228)	(0.229)	(0.229)
DT214	-0.720***	-0.730***	-0.754***	-0.761***	-0.767***	-0.768***
	(0.267)	(0.268)	(0.269)	(0.269)	(0.269)	(0.268)
DT236	0.125	0.107	0.109	0.110	0.109	0.109
	(0.112)	(0.112)	(0.112)	(0.111)	(0.111)	(0.111)
DT242	-0.415***	-0.425***	-0.403***	-0.393***	-0.386***	-0.387***
	(0.121)	(0.120)	(0.121)	(0.120)	(0.121)	(0.121)
DT254	-0.446***	-0.450***	-0.447***	-0.444***	-0.444***	-0.445***
	(0.160)	(0.159)	(0.159)	(0.159)	(0.160)	(0.160)
DT268	-0.508***	-0.466***	-0.426***	-0.423***	-0.432***	-0.436***
	(0.122)	(0.122)	(0.122)	(0.122)	(0.122)	(0.122)
DT292	-0.941***	-0.900***	-0.882***	-0.879***	-0.881***	-0.883***
	(0.128)	(0.130)	(0.130)	(0.130)	(0.130)	(0.130)
DT311	-0.913***	-0.830***	-0.785***	-0.788***	-0.810***	-0.813***
	(0.214)	(0.213)	(0.214)	(0.213)	(0.213)	(0.213)
DT349	0.809**	0.756**	0.736**	0.740**	0.753**	0.757**
	(0.332)	(0.327)	(0.325)	(0.324)	(0.324)	(0.325)
DT377	-1.043***	-1.040***	-1.039***	-1.041***	-1.049***	-1.050***
	(0.219)	(0.220)	(0.219)	(0.220)	(0.220)	(0.220)
DT532	-0.353**	-0.366**	-0.356**	-0.350**	-0.346**	-0.346**
	(0.146)	(0.148)	(0.149)	(0.149)	(0.149)	(0.149)
DT548	0.340***	0.327***	0.337***	0.341***	0.342***	0.341***
	(0.084)	(0.085)	(0.085)	(0.085)	(0.085)	(0.085)
DT573	0.031	0.027	0.038	0.040	0.038	0.036
	(0.233)	(0.231)	(0.232)	(0.233)	(0.235)	(0.235)
DT581	-0.454**	-0.463**	-0.460**	-0.462**	-0.470***	-0.473***
	(0.181)	(0.180)	(0.180)	(0.180)	(0.180)	(0.179)
DT615	-0.605***	-0.631***	-0.632***	-0.628***	-0.626***	-0.627***
	(0.163)	(0.164)	(0.164)	(0.164)	(0.163)	(0.163)
DT644	0.106	0.102	0.113	0.113	0.108	0.107
	(0.128)	(0.126)	(0.125)	(0.125)	(0.125)	(0.125)
DT701	-0.508***	-0.481***	-0.444***	-0.438***	-0.442***	-0.444***
	(0.127)	(0.128)	(0.128)	(0.128)	(0.128)	(0.128)
DT720	0.263	0.387**	0.349*	0.337*	0.348*	0.355*
	(0.190)	(0.181)	(0.186)	(0.188)	(0.186)	(0.185)
DT782	-0.393	-0.345	-0.321	-0.315	-0.310	-0.311
	(0.262)	(0.265)	(0.269)	(0.270)	(0.271)	(0.271)
DT869	-0.661***	-0.664***	-0.656***	-0.653***	-0.655***	-0.657***

	(0.183)	(0.183)	(0.182)	(0.181)	(0.180)	(0.180)
DT909	-0.560***	-0.528***	-0.530***	-0.537***	-0.552***	-0.556***
	(0.197)	(0.196)	(0.194)	(0.193)	(0.193)	(0.193)
DT919	-0.974***	-0.966***	-0.951***	-0.952***	-0.966***	-0.971***
	(0.274)	(0.268)	(0.261)	(0.260)	(0.261)	(0.262)
DT921	-0.659***	-0.641***	-0.656***	-0.657***	-0.652***	-0.653***
	(0.204)	(0.207)	(0.206)	(0.206)	(0.205)	(0.205)
DT923	-0.244	-0.270	-0.266	-0.261	-0.258	-0.260
	(0.212)	(0.213)	(0.214)	(0.214)	(0.214)	(0.214)
DT926	-0.015	-0.013	-0.009	-0.008	-0.011	-0.013
	(0.153)	(0.153)	(0.153)	(0.153)	(0.154)	(0.154)
DT940	-0.940***	-0.967***	-0.952***	-0.943***	-0.934***	-0.933***
	(0.207)	(0.206)	(0.207)	(0.208)	(0.208)	(0.208)
DT949	-0.081	-0.100	-0.102	-0.098	-0.092	-0.092
	(0.171)	(0.173)	(0.174)	(0.174)	(0.174)	(0.174)
DT963	-0.251	-0.296*	-0.311*	-0.311*	-0.308*	-0.307*
	(0.178)	(0.179)	(0.179)	(0.179)	(0.178)	(0.178)
DT976	-0.463*	-0.396	-0.418	-0.422*	-0.420*	-0.424*
	(0.253)	(0.259)	(0.255)	(0.254)	(0.253)	(0.253)
DT987	0.214	0.262	0.287	0.294	0.298	0.297
	(0.260)	(0.266)	(0.271)	(0.272)	(0.274)	(0.275)
DT994	-0.814***	-0.798***	-0.813***	-0.817***	-0.820***	-0.820***
	(0.244)	(0.246)	(0.248)	(0.248)	(0.249)	(0.249)
DT1010	-0.778***	-0.739***	-0.738***	-0.739***	-0.738***	-0.738***
	(0.209)	(0.212)	(0.214)	(0.215)	(0.215)	(0.215)
DT1011	-0.793***	-0.820***	-0.785***	-0.770***	-0.762***	-0.763***
	(0.279)	(0.279)	(0.276)	(0.275)	(0.275)	(0.275)
DT1020	-0.234	-0.261	-0.278	-0.279	-0.276	-0.275
	(0.240)	(0.237)	(0.238)	(0.238)	(0.239)	(0.239)
DT1071	0.129*	0.098	0.102	0.110	0.118*	0.118*
	(0.066)	(0.067)	(0.067)	(0.067)	(0.067)	(0.067)
DT1128	-0.178	-0.203	-0.202	-0.202	-0.205	-0.205
	(0.134)	(0.134)	(0.133)	(0.133)	(0.132)	(0.132)
DT1222	0.114	0.116	0.130	0.135	0.138	0.137
	(0.088)	(0.089)	(0.090)	(0.090)	(0.090)	(0.090)
DT1227	0.254	0.275	0.315	0.318	0.306	0.303
	(0.310)	(0.311)	(0.310)	(0.309)	(0.309)	(0.310)
DT1230	-0.040	-0.032	-0.048	-0.054	-0.061	-0.063
	(0.209)	(0.211)	(0.211)	(0.211)	(0.210)	(0.210)
DT1245	-0.406**	-0.405**	-0.403*	-0.402*	-0.403*	-0.404*
	(0.204)	(0.205)	(0.206)	(0.207)	(0.208)	(0.208)
DT1253	-0.223*	-0.217	-0.196	-0.189	-0.183	-0.184
	(0.133)	(0.134)	(0.134)	(0.134)	(0.135)	(0.135)
DT1303	0.101	0.065	0.047	0.046	0.049	0.049
	(0.251)	(0.255)	(0.255)	(0.254)	(0.252)	(0.251)
DT1342	-0.310	-0.339*	-0.366*	-0.371*	-0.372*	-0.371*
	(0.194)	(0.197)	(0.199)	(0.199)	(0.199)	(0.199)
DT10006	-1.059***	-1.025***	-1.020***	-1.019***	-1.017***	-1.017***
	(0.260)	(0.266)	(0.269)	(0.269)	(0.269)	(0.269)
DT10026	-0.153	-0.126	-0.062	-0.048	-0.054	-0.061
	(0.332)	(0.330)	(0.324)	(0.323)	(0.325)	(0.326)
DT10027	0.010	0.063	0.080	0.079	0.068	0.063
	(0.291)	(0.288)	(0.285)	(0.285)	(0.285)	(0.285)
DT10028	-0.735***	-0.690***	-0.665**	-0.669**	-0.689***	-0.694***
	(0.268)	(0.267)	(0.264)	(0.263)	(0.261)	(0.262)

DT10030	-1.082***	-1.099***	-1.114***	-1.117***	-1.121***	-1.123***
	(0.276)	(0.277)	(0.278)	(0.279)	(0.279)	(0.279)
DT10031	-0.631**	-0.671**	-0.701**	-0.704**	-0.697**	-0.693**
	(0.291)	(0.300)	(0.303)	(0.303)	(0.302)	(0.301)
DT10035	-0.411*	-0.416*	-0.409*	-0.407*	-0.410*	-0.413*
	(0.242)	(0.242)	(0.243)	(0.243)	(0.243)	(0.243)
DT10037	-0.462***	-0.467***	-0.463***	-0.463***	-0.467***	-0.468***
	(0.170)	(0.172)	(0.173)	(0.173)	(0.173)	(0.172)
DT10038	-1.266***	-1.301***	-1.287***	-1.277***	-1.265***	-1.266***
	(0.231)	(0.233)	(0.232)	(0.231)	(0.230)	(0.230)
DT10040	-0.452***	-0.445***	-0.458***	-0.462***	-0.465***	-0.465***
	(0.164)	(0.165)	(0.166)	(0.166)	(0.167)	(0.167)
DT10041	-0.597***	-0.609***	-0.611***	-0.611***	-0.612***	-0.613***
	(0.228)	(0.229)	(0.229)	(0.229)	(0.229)	(0.229)
DT10042	-0.694***	-0.681***	-0.668***	-0.666***	-0.665***	-0.664***
	(0.190)	(0.193)	(0.193)	(0.192)	(0.191)	(0.191)
DT10043	-0.178	-0.194	-0.205	-0.206	-0.210	-0.212
	(0.225)	(0.226)	(0.226)	(0.226)	(0.226)	(0.226)
DT10046	0.084	0.060	0.064	0.068	0.072	0.071
	(0.327)	(0.328)	(0.327)	(0.326)	(0.325)	(0.324)
DT10048	0.042	0.046	0.042	0.039	0.032	0.030
	(0.233)	(0.234)	(0.234)	(0.234)	(0.234)	(0.234)
DT10049	-0.191	-0.206	-0.215	-0.217	-0.221	-0.222
	(0.167)	(0.168)	(0.168)	(0.167)	(0.167)	(0.167)
DT10050	-0.375	-0.386	-0.390*	-0.387	-0.379	-0.377
	(0.231)	(0.235)	(0.236)	(0.236)	(0.236)	(0.236)
Constant	0.091***	-0.320***	-0.638***	-0.754***	-0.929***	-0.981***
	(0.031)	(0.045)	(0.065)	(0.057)	(0.067)	(0.073)
Observations	33,258	33,258	33,258	33,258	33,258	33,258

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1