HEDG

HEALTH, ECONOMETRICS AND DATA GROUP

THE UNIVERSITY of York

WP 23/12

The Value of Patients: Heterogenous Physician Learning and Generic Drug Diffusion

Zhengnan Zhu

August 2023

http://www.york.ac.uk/economics/postgrad/herc/hedg/wps/

The Value of Patients: Heterogenous Physician Learning and Generic Drug Diffusion

Zhengnan Zhu
Imperial College London

PRELIMINARY DRAFT

May 2023

Abstract

This paper explores how the difference in the quantity and quality of information received by physicians shapes the learning process and subsequently the diffusion of generic drugs. By exploiting prescription level data, I find that both the volume of information and the difference in the composition of information signals received by a physician contributes to the heterogeneity in adoption rates. In particular, having more information signals from new patients who move from peers increases the adoption rate of generic drugs. To explain the findings, I develop a physician learning framework where the informativeness of signals differ across old patients and new patients from other doctors. The calibrated results suggest that new patient signals weigh more than own patient signals in directly raising physicians' expectations on the true quality, whilst this effect does not act through reducing uncertainty around the expectation. The results on the compositional effect of information echoes with "the strength of weak ties" where new patients from peers, seen as weak ties, are more informative in raising physicians' optimism of new drugs.

Keywords: learning, information, diffusion processes, network
JEL Classification: O33, D83, D85

1 Introduction

Understanding the diffusion process of new ideas, products, or technologies is of great interest to economists because it can have significant economic implications. The aggregate diffusion process is determined by individuals' adoption decisions, and the heterogeneity across individuals can result in difference in adoption time and volume. Individuals' adoption decisions can be influenced by factors such as quality difference of the innovation, perceptions about the innovation, and/or social influence. Among the diffusion models

that focus on adoption agents' decisions[1], social learning models are widely applicable in most scenarios, where agents extract and make rational use of information to help form belief on the value (e.g. quality, effectiveness) of the innovation and reach a decision (Young, 2009). The differences in the access to and quality of the information result in different belief formation which may further explain differences in adoptions. Micro-level studies have investigated the role and channels of learning on technology diffusion in many different economic contexts. For instance, learning from own and neighbours experience reduces the knowledge barrier to the new agricultural technology and increases its adoption rate (Foster and Rosenzweig, 1995; Bandiera and Rasul, 2006; Conley and Udry, 2010).

Learning is particularly important in the healthcare context. Doctors have habits and inertia in treatment decisions developed throughout medical practice. When a new medical treatment is on the market, the uncertainty about the effectiveness/adequacy in general and for a specific patient requires doctors to obtain enough information to switch and adopt. Doctors can form their perceptions about medical treatment from direct experiences (when she gives the innovation to her patients) (Ching, 2010; Coscelli and Shum, 2004; Crawford and Shum, 2005) and/or other sources of information (e.g. from guidelines, advertising, medical conferences, and peers experience) (Erdem et al., 2008; Epstein and Ketcham, 2014; Arrow et al., 2020). Empirical studies have examined the effect of different information sources on doctors' decision makings where some of them explicitly model the learning process and analyze how individuals react to information signals (Crawford and Shum, 2005; Narayanan and Manchanda, 2009; Ching and Ishihara, 2010, 2012, e.g.) and others look for statistical relevance between information sources and decision-makings (Mizik and Jacobson, 2004; Epstein and Ketcham, 2014; Arrow et al., 2020).

The aim of this paper is to shed more light on this topic by examining how the difference in the quantity and quality of information from patients received by physicians shapes the learning process and subsequently the innovation diffusion patterns. Using the universe of Atorvastatins prescriptions in Finland, I study the adoption of generic drugs in the population from 2008 to 2011 in which the generic version first came into existence. Since generic drugs are bioequivalent to branded drugs, looking at the adoption of generic drugs abstracts away the concern of inherent quality difference, which can be a fundamental determinant of the innovation diffusion process. The lags in physicians' adoptions of generic drugs are due to perceived quality difference between branded and generic drugs. The empirical evidence shows that not only the volume of information on

[1] Young (2009) discusses three broad classes of innovation diffusion models – contagion, social influence, social learning – that considers different diffusion mechanisms driven by adoption agents. Other diffusion models consider external factors (e.g. price, quality change) or the combinations between the two (Bass, 1969, e.g.)

the drug experience matters for physicians' learning and adoption of the generic drugs, but the sources of information also matter. Conditional on the same level of information received from patients, higher share of information signals received from new patients who come from other doctors increases the adoption rate of generic drugs. However, although greater volume of information on drug experience reduces the variation in adoption rates, having higher new signal shares has a counter effect.

The distinctive informational effects between different signals echoes with "the strength of weak ties" theory in social networks, which is first put forward by Granovetter (1973). The theory posits that people involved in frequent interactions tend to have repeated and/or overlapping information, whilst infrequent relationships, known as "weak ties", are more beneficial in terms of information transmission. Empirical studies document that weak ties are useful in the job market (Granovetter, 1973; Rajkumar et al., 2022), technological information diffusion, and immigration. In the context of learning from patients' experience, a doctor's new patients moving from other doctors are seen as weaker ties comparing to a doctor's old patients, who have had at least one interaction with the doctor before the arrival of new patients. Albeit less frequent interactions, new patients who recently move from other doctors bring novel experiences on the drug, in contrast to old patients who convey repeated information. For example, new patients can bring experiences on generic drugs combined with own characteristics, which are new to the doctor. Positive experience with generic drugs from elsewhere also provides validating information that raises the optimism about the drug for the doctor.

Although research often connects social networks, spillover effects, and learning in various settings, there is limited evidence on the specific discussions of how learning acts among weak ties. Granovetter (1973) documents the correlation that a large proportion of jobs were obtained through "weak ties". By conducting randomized experiments on Linkedin, Rajkumar et al. (2022) show that moderately weak ties are the most beneficial in providing job opportunities. However, Jackson et al. (2008) raises concerns that are not fully answered in the literature. Firstly, the benefit from weak ties could arise from the fact that individuals tend to have many more weak ties than strong ties, hence the usefulness of weak ties could be a scale effect. Secondly, weak ties might be intrinsically different from strong ties, regardless of the bridging behaviours. Thirdly, it is questionable whether strong ties indeed have overlapping information in their neighbourhoods. Hence, a careful exploration on the interaction structures is important to examine the role of weak ties.

To address the important questions and the empirical findings from data, I develop a Bayesian learning framework similar to ones in Coscelli and Shum (2004); Crawford and Shum (2005); Narayanan and Manchanda (2009), where physicians learn from patients' experiences to update their belief about the quality of drugs. When the decision process entails repeated decision makings, it makes sense for individuals to gradually form the

belief by considering all past actions and assign appropriate weights to the experience, which is the essence of the Bayesian learning process and different from other learning types.[2] Ching (2010) models a representative physician who aggregates information from patients and the learning is homogeneous across patient types (two types in estimation). The heterogeneity in demand lies in different price sensitivities of patients. Using the same learning framework, Ching and Ishihara (2012) examine the effect of detailing by introducing the probability that physicians are either informed or uninformed, determined by the detailing stock, which subsequently affects the market demand. These two papers predict total adoption/demand after learning and match product sales data for a number of generic drugs (Ching, 2010) and ACE inhibitors (Ching and Ishihara, 2012). Coscelli and Shum (2004) look at the diffusion of a new anti-ulcer drug which can treat more than one symptom. The model allows spillover effects where the information signals are correlated across symptoms, and physicians know the correlation structure. Although a physician only observes one signal from one condition each time she sees a patient from the associated diagnosis group, she updates the belief across all attributes of the drug. Narayanan and Manchanda (2009) consider the difference in physicians' learning rates by modelling physician-specific updating variances and use physician prescription sequence without identifying patients. Given the same information signals, a larger signal variance slows down the speed of learning. They consider both patients experience and market communications (detailing) and confirm that the responsiveness to signals differ across physicians and over time. They then draw insights on how pharmaceutical firms can improve the targeting of detailing across physicians. Similar in the learning manner but with a different emphasis, Crawford and Shum (2005) address patient-drug match values to explain the observed drug choices and treatment lengths of patients. They model learning across symptomatic and curative effects for multiple drugs and heterogeneity in drug effectiveness across patients. By observing the sequence of patient visits, they find that learning allows consumers to reduce uncertainty.

I contribute to the literature in several aspects. Firstly, I distinguish information signals brought by old VS new patients for a given physician at a given time. To do so, I incorporate patient movements between physicians over time in the model, which generate the variations in the composition of information signals across physicians. The value of patients in contributing to the heterogeneity in adoption rates lies in the compositional effect of information, and not in patient-specific characteristics such as price sensitivity (Ching, 2010), diagnosis groups (Coscelli and Shum, 2004), or drug effectiveness (Crawford and Shum, 2005). The physician-patient prescription level dataset facilitates the analyses whilst aggregate sales data (Ching, 2010) and sequence of visits of patients without physician prescriptions (Crawford and Shum, 2005) are not able to capture the

[2]There are also other learning types such as "rules of thumb" learning (Ellison and Fudenberg, 1993, 1995; Banerjee, 1992), "persuasion bias" (DeMarzo et al., 2003), Markov process (DeGroot, 1974).

dynamics of information flow in the population and the changes in physicians' adoption behaviours over time. Secondly, I add to observational studies on the strength of weak ties by explicitly modelling different learning effects from strong VS weak ties. Since I observe the proportion of new patient signals relative to total number of patient signals, I can control for potential scale effects of weak ties. In addition, patients as information signals only differ in the composition for each physician. Hence it is not the intrinsic characteristics or bridging behaviours of a specific patient that alter the learning rate of the focal physician. A physician's old patients are defined to have overlapped or repeated information comparing to new patients, since old patients have had interactions with the physician.

Two sources of heterogeneity in information are generated from the model. The first difference is the number of signals (volume of information) and the second is the difference in the share of signals by old VS new patients. The model is then calibrated to fit the overall generic diffusion patterns in the population. The learning process is quicker if a physician has a higher total number of signals. In addition, the calibrated model draws further insights on the learning mechanisms induced by information signals in a Bayesian learning framework. Narayanan and Manchanda (2009) assume signal variances are physician-specific, which results in speed difference in belief updating. Coscelli and Shum (2004) show that although physicians are initially confident in pessimistic priors, they update quickly due to precise signals (estimated variances are small). In my framework, I separately identify the effect of learning on optimism and reduction in uncertainty. The calibrated model implies that the effect of new patient information signals act through directly raising physicians' expectations on the quality of generic drugs, whilst not reducing the variation (uncertainty) in the quality expectations.

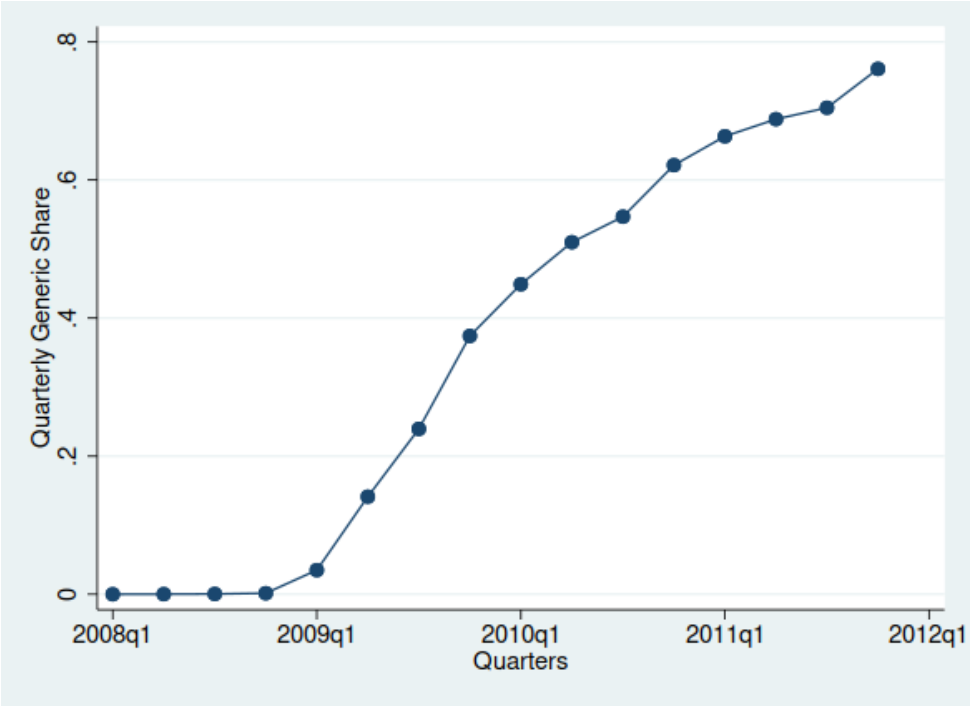
The rest of the paper is organized as follows. Section 2 describes the dataset and presents empirical evidence, which motivates the learning framework developed in Section 3. Section 4 presents the calibration details. Section 5 discusses the implications of estimation results. Section 6 conducts counterfactural simulations. Section 7 concludes.

2 Background and Data Patterns

2.1 Institutional Settings

In Finland, primary health care (family doctor) services are provided by municipal health centres, and specialised medical care is provided by hospital-district hospitals, which are funded by member municipalities. At primary care level, patients can choose between municipal health care, private health care, and occupational health care, while most patients use municipal health care. Under municipal health care coverage, patients are treated in the municipal health centre where she is registered, based on her residency

Figure 1: Quarterly Adoption Rate of Generic Atorvastatins in the Population



address. Until 2011, physician choice was limited with patients assigned to physicians based on residency or employer (Mossialos and Srivastava, 2008; Keskimaki et al., 2019). The choice of secondary care (hospital care) providers is also limited, with patients referred by GPs to the district hospital in the immediacy of her residency place.

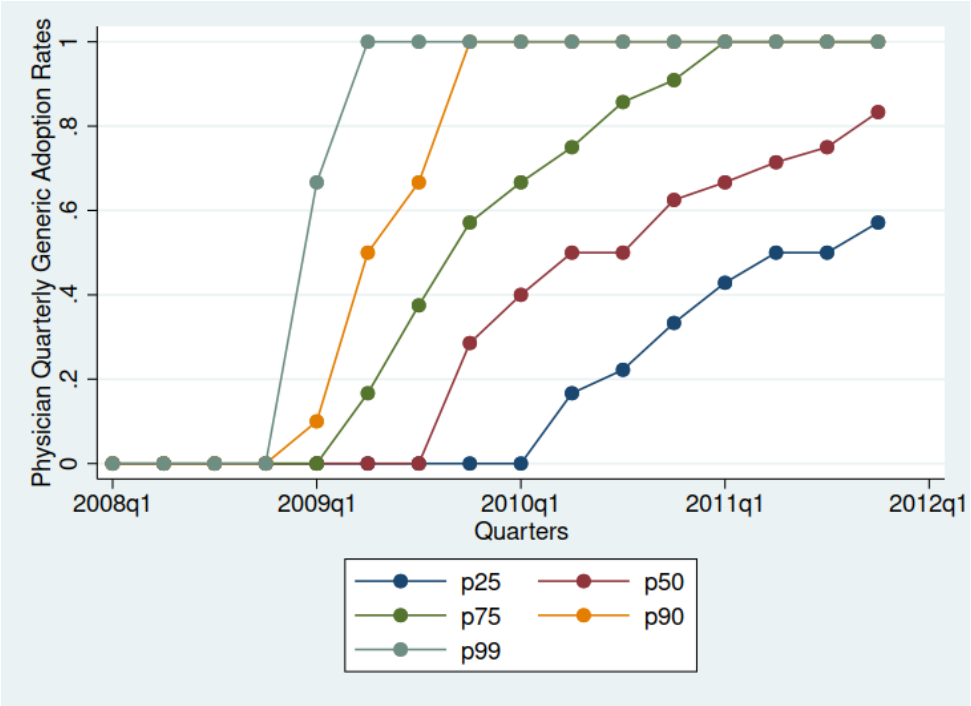
Physicians in Finland work for only one care provider in the public sector and receive a combination of payments, including basic salaries and payment based on the total volume of patients they see (Keskimaki et al., 2019). Therefore, physicians' earnings are not affected by the specific treatment provided including the particular drugs prescribed to patients.

Patients buy prescription drugs at pharmacies. FIMEA (The Finnish Medicines Agency), the regulation authority of pharmaceuticals in Finland, decides the wholesale and retail prices of pharmaceuticals (WHO, 2019). As a result, the pharmaceutical pricing is uniform across pharmacies. Pharmaceutical companies can apply to change the wholesale prices every 2 weeks.

2.2 Data Descriptions

The data, obtained from the Social Insurance Institution in Finland - KELA, contains the universe of Atorvastatin prescriptions for Finnish population. Atorvastatin is a type of statin and falls in Anatomical Therapeutic Class (ATC) C10AA05. Patients diagnosed to take statins need to receive repeated prescriptions over time since statins are used to lower

Figure 2: Physician Quarterly Adoption Rate by Percentiles



the cholesterol level in the blood to prevent diseases such as heart attacks or strokes. I consider the time periods from 2008 to 2011 since the first generic version of Atorvastatin went onto the Finnish market in January 2008. Each observation is defined at the level of a prescription, where I observe the prescribed drug and its version (branded or generic), the physician who prescribed the drug and the patient to whom the drug was prescribed, and the prescription date. Subsequently, it allows me to track prescription histories and patients' movements between physicians over time. The data is then aggregated to a quarterly level in terms of time to observe the overall generic adoptions in the population.

Figure 1 shows the overall trend of quarterly generic adoption rates of Atorvastatins in the population. The generic adoption takes off one year after the initial launch, with a rapid increase during the subsequent 7-8 quarters up to more than 60%, and grows slowly afterwards. Although generic drugs are bioequivalent to branded drugs, data shows that physicians do not immediately switch to generics. Figure 2 plots physicians' quarterly generic adoption rates in the population by percentiles. When the adoption of generics starts in the first quarter of 2009, the 99th percentile of physicians' generic adoption rates is around 65%, while 90% of the adoption rates are below 10%. The gaps between percentiles are the largest in the early stage of the diffusion, and gradually decrease as generics diffuse in the population. Such data patterns in the two figures are consistent with the idea that physicians gradually update their *perceived* quality gap between branded and generic drugs, and the heterogeneity across physicians is present in the population.

Table 1: The Relationship Between Information signals and Generic Adoptions

Variables	(1) $genshare_{jt}$	(2) $genshare_{jt}$
ln(total signal no.)	0.108***	0.105***
	(0.006)	(0.006)
New signal share		0.178***
		(0.059)
Physician Fixed Effects	Yes	Yes
Observations	14,712	14,712
R-squared	0.311	0.311

2.3 Reduced Form Evidence

One important source of physician learning is "gains from trial information" (Erdem and Keane, 1996; Ching et al., 2013), i.e. from patients' experiences on the prescriptions. In this context, I define a patient as an information signal if she has used a generic drug once. If a physician j prescribes a generic drug to an existing patient, this patient counts as an old-patient signal for this physician; if a physician j has a new patient moving from another physician and this new patient was prescribed a generic drug before visiting j, she counts as a new-patient signal for this physician.

To demonstrate how patients' information on the experience of generic drugs shapes physician learning and adoption decisions, I keep physicians who prescribe in every quarter from 2009 to 2011 (since the generic adoption starts from 2009) and run the following two regressions:

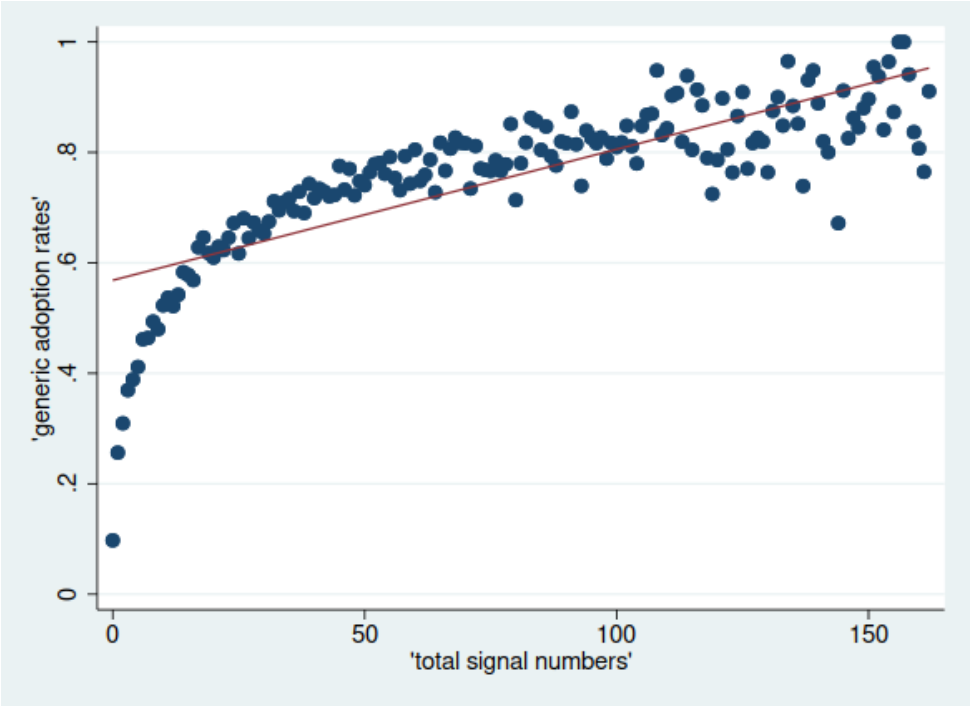
$$genshare_{jt} = \beta_0 + \beta_1 ln(totalsignalnumbers)_{jt} + \eta_j + \epsilon_{jt} \quad (1)$$

$$genshare_{jt} = \beta_0 + \beta_1 ln(totalsignalnumbers)_{jt} + \beta_2 newsignalshare_{jt} + \eta_j + \epsilon_{jt} \quad (2)$$

where the dependent variable is the generic share of a physician j in quarter t, denoted as $genshare_{jt}$. *Total signal numbers* that a physician j has in quarter t is the sum of old-patient signals and new-patient signals that j accumulates ***before*** t. I then calculate the share of new patient signals for each physician given the total signals numbers for each quarter t, denoted as $newsignalshare_{jt}$. η_j is physician-level fixed effects to control for unobserved physician characteristics that may be correlated with their prescription behaviours. The results in Table 1 show that the quarterly share of generic adoptions for each physician is positively correlated with the total number of signals accumulated before t. Conditional on the same number of total signals, having more new patient signals further increases generic adoption rates.

In addition to correlation patterns on individual levels, I further look at the aggregate relationship between the total signal numbers and generic adoption rates in the population. In figure 3, each dot represents the average generic adoption rates calculated within

Figure 3: Generic Adoption Rates by Signal Numbers



a group of physicians defined by the same total signal numbers, regardless of time. The positive relationship shows that more signals are accumulated, higher the average generic adoption rate is.

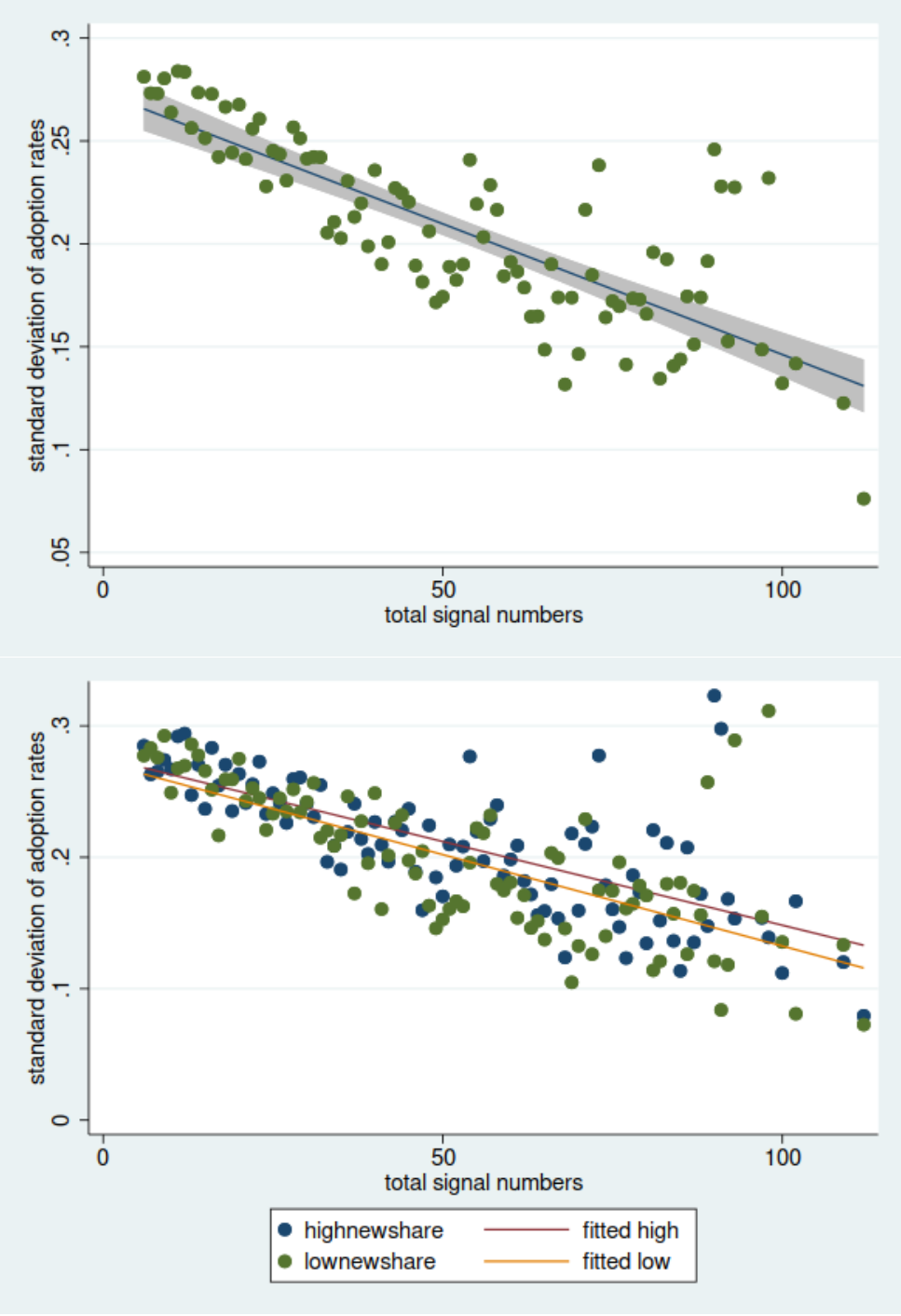
I further examine how information signals affect the variation of generic adoption rates by running the following regressions:

$$sd(adoption)_s = \alpha_0 + \alpha_1 ln(totalsignalnumbers)_s + \epsilon_s \tag{3}$$

$$sd(adoption)_{sg} = \alpha_0 + \alpha_1 ln(totalsignalnumbers)_s + \alpha_2 \mathbb{1}(highnewsignalshare)_{sg} + \epsilon_s \tag{4}$$

where the dependent variable in equation 3 is the standard deviation of adoption rates across physicians with the same total number of signals (regardless of quarters). The observations are collapsed on the level of signal numbers in equation 3, using subscript s. Within each signal number group s, I rank physicians' shares of new patient signals. If a physician's new signals share is above the median of new signals shares within the same total signal number group, she is considered in the *highnewsignalshare* group. Otherwise, she is considered in the *lownewsignalshare* group. By doing this, I further divide physicians into two subgroups by their share of new patient signals within each total signal number group, using subscript sg in equation 4. I then recalculate the standard deviation of adoption rates within *highnewsignalshare* group and *lownewsignalshare* group respectively, conditional on the same total signal numbers. $\mathbb{1}(highnewsignalshare)_{sg}$ is an indicator variable, which is equal to one if the group is considered to have high new signal shares

Figure 4: Variation of Adoption Rates Within Groups by Signal Numbers



Notes: In the top panel, each dot represents the standard deviation of adoption rates calculated within a group defined by the same total signal numbers. In the bottom panel, I divide physicians into two subgroups according to whether they have new patient signal shares above or below the median, within each total signal number group. The standard deviations are calculated within each subgroup. I count the number of observations in each subgroup and only keep the ones that have observations more than the median (6 observations). After removing outliers, unbalanced cases where only one subgroup is left are also omitted.

Table 2: The Relationship between Signals and Variation in Adoption Rates

Variables	(1) $sd(adoption)$	(2) $sd(adoption)$
ln(total signal no.)	-0.051***	-0.053***
	(0.004)	(0.004)
Group(high new share)		0.010
		(0.005)
Observations	94	188
R-squared	0.656	0.517

within the same total signal number group. I count the number of observations in each subgroup across all signal number groups and omit the ones that have observations less than the median (6 observations). To have a balanced sample structure, I only keep the observations where both subgroups exist after the sample cut for the same total signal numbers. This results in 94 signal number groups and 188 subgroups, when split by the share of new patient signals.

The top panel of figure 4 shows a negative relationship between the total number of signals and the standard deviation of adoption rates within each signal number group. This indicates that as physicians accumulate more experience information from patients, the variations around their adoption behaviours become smaller. Studies suggest that the uncertainty of drug qualities can be reduced by learning more information from experiences (Crawford and Shum, 2005). In the bottom panel of figure 4, physicians are split into two subgroups within the same signal number group: those with high new signals share and those with low new signals share. The negative relationship persists and the overall level of the standard deviation for *highnewsignalshare* group is slightly higher. The regression results for equation 3 and 4 are presented in Table 2 to quantify the magnitudes. Having a higher share of new patient signals do not appear to be statistically significant in affecting the standard deviation of adoption rates.

3 Model

To understand the role of learning in driving the adoption patterns from data, I develop a model that characterizes how physicians learn from patients' experiences. The sequence of the events is the following. At the beginning of time period t, a physician j decides on behalf of her patients based on information set I_{jt}, which contains the physician's beliefs about the quality of drugs. Patients receive branded (b) or generic (g) drugs and reveal their experiences to physicians. At the beginning of $t-1$, physicians aggregate information from patients' experience in t and update their beliefs about the quality of generic drugs, which constitute the information set that physicians use in the beginning

of the next period: $I_{j,t+1}$.

A patient i's utility of taking drug $d \in \{b, g\}$ at time t takes the form:

$$u_{idt} = a_{idt} - c_i p_{dt} + z_{idt} \tag{5}$$

a_{idt} is i's actual experience of consuming drug $d \in \{b, g\}$ at time t and given by:

$$a_{idt} = \theta_d + \varepsilon_{idt} \tag{6}$$

where θ_d is the mean quality of a drug and $\varepsilon_{idt} \sim N(0, \sigma_\varepsilon^2)$. c_i is the copayment rate of patient i, which is solely determined by the severity of the disease. p_{dt} is the unit price (package price/Defined Daily Doses (DDD)) of drug d, which is exogenously given when making prescription decisions, since prices of drugs are regulated by the authority rather than endogenously determined by drug choices. z_{idt} is a patient-drug specific shock that is observable to physicians at the time of prescription but unobservable to econometricians, which captures the patient-specific uncertainty about a drug and $z_{idt} \sim N(0, \sigma_z^2)$.

A physician is assumed to act as a perfect agent for patients in the decision-making process. Conditional on her information set I_{jt}, physician j will prescribe a generic drug g to patient i at time t if $E[u_{igt}|I_{jt}] > E[u_{ibt}|I_{jt}]$. At time t, physician j maximizes expected utility:

$$\begin{aligned} E[U_{jt}|I_{jt}] &= \int_i max\{E[u_{igt}|I_{jt}], E[u_{ibt}|I_{jt}]\}di \\ &= \int_{i \in s_{jgt}} (E_{jt}[\theta_g] - c_i p_{gt} + z_{igt})di + \int_{i \in s_{jbt}} (\theta_b - c_i p_{bt} + z_{ibt})di \end{aligned} \tag{7}$$

where s_{jgt} (s_{jbt}) is the set of patients that receive generic (branded) prescriptions from physician j at time t. The mean quality of branded drugs (θ_b) is assumed to be known to the public since branded drugs have been on the market for a long time when generic drugs come onto the market. As illustrated by data, the adoptions do not immediately take place when generic versions come into existence. Physicians have *perceived* quality differences that hinder generic adoptions and they learn about the mean quality of generic drugs, θ_g, from patients' experiences.

3.1 Learning

The initial belief about θ_g is given by $N(\bar{\theta}, \sigma_\theta^2)$, $\bar{\theta} < \theta_b$. At time t, physician j's belief on θ_g is $\theta_g \sim N(\mu_{jgt}, v_{jgt}^2)$, where μ_{jgt} and v_{jgt}^2 are updated according to Bayes' rule:

$$\mu_{jgt} = \frac{v_{jgt}^2}{\sigma_\theta^2}\bar{\theta} + \frac{v_{jgt}^2}{\sigma_\varepsilon^2} S_{jgt} \bar{a}_{jgt} \tag{8}$$

$$v_{jgt}^2 = \frac{\sigma_\theta^2 \sigma_\varepsilon^2}{\sigma_\varepsilon^2 + S_{jgt}\sigma_\theta^2} \tag{9}$$

If a patient i has been prescribed a generic drug, she counts as a "signal" to a physician ($i \in s_{jgt}$), which means that the patient conveys information on the experience of generic drugs (a_{igt}). $S_{jgt} = \sum_{\tau=1}^{t-1} s_{jg\tau}$ is the total number of signals physician j receives **up to time** t, $\bar{a}_{jgt} = (\sum_{\tau=1}^{t-1} \int_{i \in s_{jg\tau}} a_{ig\tau} di) / \sum_{\tau=1}^{t-1} s_{jg\tau}$ is the mean signals physician j receives **up to time** t. Ceteris paribus, equation 9 suggests that v_{jgt}^2 will be lower if: a) experience variability σ_ε^2 is lower; b) physicians have higher number of signals.

3.2 Patients' Movement

As explained in the institutional background, individuals can only choose healthcare services within a limited range in the residence area. A patient changes doctors if she moves residence address or is referred to a hospital specialist by a GP in the same residence area. It is plausible to assume that patients' movement is exogenous to the prescription choice of physicians. Physicians do not have the incentives to attract patients through prescriptions. In the model, physicians take incoming patients as exogenous information signals and maximize per period utility for each patient.

Assume the total number patients in the population is N[3] and the number of patients in each period is N_t. $N_t \leq N$ since not necessarily every patient sees a physician at time t. Conditional on a visit to a physician at $t-1$, patients will continue to see a physician at t with probability λ_t^{cont}. Patients who did not see a physician at $t-1$ will see a physcian at t with probability λ_t^{rest}. Therefore, active patients in each period t are given by:

$$N_t = N_{t-1}\lambda_t^{cont} + (N - N_{t-1})\lambda_t^{rest} \tag{10}$$

Conditional on consecutive visits in two time periods, a patient i who visits physician j at time $t-1$ may stay with j at t or move to another physician with probability p. Given the distribution of patients in the initial period $t = 1$, the total number of patients of each physician j at time t (N_{jt}) is given by:

$$N_{jt} = \lambda_t^{cont}\left[(1-p)N_{jt-1} + \frac{N_{j1}}{N_1}\sum_{j' \neq j} pN_{j't-1}\right] + \frac{N_{j1}}{N_1}\lambda_t^{rest}(N - N_{t-1}) \tag{11}$$

where N_{j1}/N_1 is the ratio of physician j's patients to all the patients in the initial period. The equation indicates that physician j has patients from three sources at time t: i) the continuing patients who visited physician j at $t-1$, ii) the continuing patients who did

[3]The entry and exit of patients and physicians to the population is not modelled in this framework since it is beyond the focus of this paper. This simplification does not affect the conclusions on learning and adoption.

not visit j at $t-1$ but move to physician j at t, and iii) patients who did not see a physician at $t-1$. For the moving patients and patients who have consultation gaps, I assume that a fraction of N_{j1}/N_1 patients will see physician j. A higher ratio indicates that the physician is more likely to get a higher volume of patients comparing to others[4] and therefore a random patient is more likely to see j.

3.3 Signal Composition

For a physician j at time t, the total number of information signals on the use of generic drugs received is $s_{jgt} = s^w_{jgt} + s^l_{jgt}$, where s^w_{jgt} and s^l_{jgt} are signals from new and old patients respectively. A physician j updates her belief at the beginning of time period t, using old patient signals of generic prescriptions in $t-1$, and new patient signals who arrive in t but received generic prescriptions in $t-1$ from other physicians. If a patient i is prescribed a generic drug in $t-1$ and moves at the end of period $t-1$, she counts as an information signal twice: an old signal for physician j' who prescribes the generic drug at $t-1$, and a new signal for physician j that patient i moves to see at t. Both j and j' incorporate the experience signal at the start of period t to form their information sets I_{jt}, $I_{j't}$ and update their beliefs according to the information sets before the decision making in period t.

To distinguish the effect of different signals, the experiences of taking generic drugs for new and old patients are rewritten respectively:

$$a^w_{igt} = \theta_g + \varepsilon^w_{igt}, \qquad \varepsilon^w_{igt} \sim N(0, \sigma^2_w) \tag{12}$$

$$a^l_{igt} = \theta_g + \varepsilon^l_{igt}, \qquad \varepsilon^l_{igt} \sim N(0, \sigma^2_l) \tag{13}$$

The two sources of signals can affect the learning process by directly raising the mean and/or reducing updating variances.

Uncertainty Channel Suppose $\sigma_w \neq \sigma_l$ (without loss of generality assume $\sigma_w < \sigma_l$). Physicians' updating on $\theta_g \sim N(\mu_{jgt}, v^2_{jgt})$ is therefore:

$$\mu_{jgt} = \frac{v^2_{jgt}}{\sigma^2_\theta}\bar{\theta} + \frac{v^2_{jgt}}{\sigma^2_w} S^w_{jgt}\bar{a}^w_{jgt} + \frac{v^2_{jgt}}{\sigma^2_l} S^l_{jgt}\bar{a}^l_{jgt} \tag{14}$$

$$v^2_{jgt} = \frac{1}{\frac{1}{\sigma^2_\theta} + \frac{S^w_{jgt}}{\sigma^2_w} + \frac{S^l_{jgt}}{\sigma^2_l}} \tag{15}$$

where S^w_{jgt} and S^l_{jgt} are the total number of signals physician j receives from new patients and old patients respectively, up to time t. $S^w_{jgt} + S^l_{jgt} = S_{jgt}$. $\bar{a}^w_{jgt}$ and $\bar{a}^l_{jgt}$ are the mean values of signals physician j receives up to time t. Given a fixed total signal number

[4]This could be due to the fact that some physicians are located in more populated areas. In addition, GPs get a higher volume of patients comparing to hospital specialists. A further examination in the data indicates that the patient share of each physician stays relatively stable over time.

S_{jgt}, equation 15 suggests that an increase in the fraction of new patient signals S^w_{jgt}/S_{jgt} will directly decrease v^2_{jgt} if $\sigma_w < \sigma_l$. Subsequently, the weight on $\bar{\theta}$ decreases due to the decrease in the variance and the weight on patient signals will become higher. Since $\bar{\theta} < \theta_b$, if realized signals ($\bar{a}^w_{jgt}$, $\bar{a}^l_{jgt}$) are greater than $\bar{\theta}$, μ_{jgt} will increase as the weights on signals increase. In other words, the increase in the fraction of new patient signals will directly decrease the variance of the mean, which subsequently contributes to faster updating of the mean by reallocating the weights to information signals.

Optimism Channel Suppose $\sigma_w = \sigma_l = \sigma_\varepsilon$ and patient signals differ in their effectiveness in raising physicians' expectation on the quality of generic drugs. Without loss of generality, new patient signals are multiplied by a factor γ in physicians' belief updating to indicate different weighting on signals:

$$\mu_{jgt} = \frac{v^2_{jgt}}{\sigma^2_\theta}\bar{\theta} + \frac{v^2_{jgt}}{\sigma^2_\varepsilon} S^w_{jgt}\bar{a}^w_{jgt} * \gamma + \frac{v^2_{jgt}}{\sigma^2_\varepsilon} S^l_{jgt}\bar{a}^l_{jgt} \tag{16}$$

and the variance of the belief is defined by equation 9. γ indicates that physicians value new patient signals differently from old patient signals. If $\bar{a}^w_{jgt} * \gamma > \bar{a}^l_{jgt}$, a higher fraction of new patient signals ($\frac{S^w_{jt}}{S_{jt}}$) will directly increase μ_{jgt} without decreasing the variance, and subsequently increase the adoption rate of generic drugs. Traditional learning models do not distinguish between the reduction of variances and the increase of means. Identifying two different channels through which physicians learn is a key contribution of my approach.

The adoption rate of generics drugs of physician j at time t is:

$$R_{jt} = \frac{\int_{i \in s_{jgt}} \mathbb{1} di}{N_{jt}} \tag{17}$$

4 Calibration and Results

The model is estimated using simulated method of moments. The comprehensive dataset allows me to determine exogenous parameters that are not affected by the decision making process and exploit the variations in data moments to calibrate parameters.

4.1 Assigned Parameters

Panel A of Table 3 reports the choices of assigned parameters. The true quality of generic drugs is assumed to be the same as the true quality of branded drugs. Without loss of generality, θ_b is normalized to zero. The copayment rate of each patient, c_i, is given according to the copayment rate distribution in the data, which is solely dependent on the severity of the disease of each patient. Patients who are diagnosed of severe morbidities have special reimbursement rate and others have basic reimbursement rate. In general, the copayment rate is 0.28 for 20% of the population and 0.58 for the rest 80%. The

Table 3: Parameterization

(a) Panel A: Assigned Parameters

Parameters	Values
θ_b	0
c_i	0.28(20%); 0.58(80%)
p_{dt}	quarterly values directly from data
T	12
$\frac{J}{N}$	1/62
λ_t^{cont}	quarterly values inferred from data
λ_t^{rest}	quarterly values inferred from data
p	0.15

(b) Panel B: Calibrated Parameters

Parameters	Descriptions	Value
$\bar{\theta}$	the initial belief on the generic quality	-0.72
θ_g	the perceived mean quality of generic drugs	-0.05
γ	optimism coefficient	-0.00142
σ_θ	s.d. of the initial belief	1.53
σ_ε	s.d. of signals	0.48
σ_z	s.d. of idiosyncratic shocks per visit	0.30

prices of generic and branded drugs are determined by the regulatory authority and given at the time of prescriptions. The unit prices are then directly taken from data and one for each quarter, since there are variations over time.

The prices of generic products are not observed in the first 4 quarters after the launch of the first generic version of Atorvastatins when there are no generic consumptions in the market. And since no generic is prescribed in the first 4 quarters, no information on the experience is available. To be consistent with the data, I set the number of periods for learning as 12 quarters. The ratio between the number of physicians and patients matches that in the data, which is 1:62.

To determine active patients in each time period, I infer from data the probabilities that describe patients' visits to physicians across time periods. λ_t^{cont} is the probability that a patient visits a physician at period t, conditional on visiting at period $t-1$. λ_t^{rest} is the probability that a patient who did not see a physician at $t-1$ sees a physician at t. Since the two probabilities vary across quarters, I calculate and assign values for each quarter. p is the moving probability of patients who visits in two consecutive quarters, which is relatively stable over time and set to 0.15.

Table 4: Goodness of Fit

Moments	Simulated	Data
M1	0.036	0.035
M2	0.134	0.149
M3	0.733	0.773
M4	0.084	0.105
M5	0.122	0.178
M6	-0.09	-0.051

4.2 Calibration

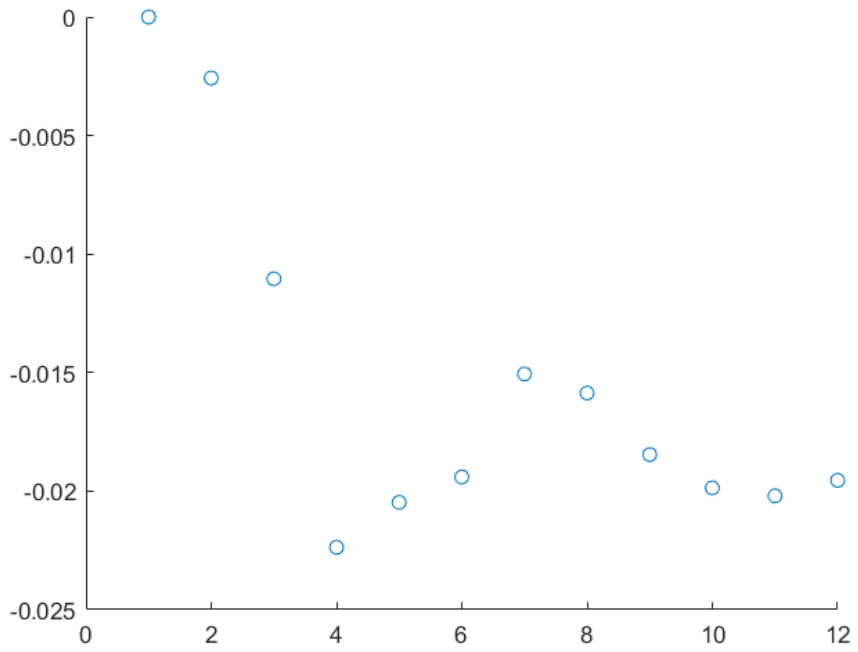
The remaining parameters are calibrated to match the overall generic diffusion patterns in the population. Specifically, I jointly estimate the parameters $\{\bar{\theta}, \theta_g, \gamma, \sigma_\theta, \sigma_\varepsilon, \sigma_z$[5]$\}$ using a set of moment conditions documented in the data. The first three moments (M1-M3) are the aggregate adoption rates at the beginning and the end of the sample period. The fourth and fifth moments (M4, M5) are the coefficient estimates of β_1 and β_2 in regression 2. The sixth moment (M6) is the coefficient estimate of α_1 in regression 3.

The calibration results are shown in Table 3 Panel B and table 4 shows the goodness of fit of the calibration results for each moment. The perceived mean quality of generic drugs is estimated to be negative (-0.05). If $\gamma > 1$, physicians are pessimistic about information brought by new patient signals and the mean updating will be slowed down. A negative estimated value of γ suggests that new patient signals directly raise physicians' expectations about the mean quality. By leveraging the universe of prescription data, the reduced form evidence shows that conditional on the same total number of signals, having a high share of new patient signals is positively correalted with generic adoption rates but does not appear to affect the variation of adoption rates. The estimation results correspond to the reduced form evidence in confirming that the compositional effect of information signals acts through the optimism channel.

Intuitively, the fact that physicians learn about the generic choices made by other physicians when receiving new patients creates a reinforced positive effect on the use of generic drugs. Without the information shocks from new patients, the possessed information will not have much variation over time since physicians can only learn from own prescriptions on the same existing patients. Consequently, the learning process will be slower. The results highlight the importance of information flow and its subsequent effect on learning and diffusion.

[5]Without loss of generality, assume $z_{ibt} = 0$ since patients have accumulated knowledge about the branded version when generic is on the market. The idiosyncratic shock associated with generic drugs per visit is captured by σ_z.

Figure 5: Percentage Differences in Adoption Rates: Absent of Signal Quality Differences

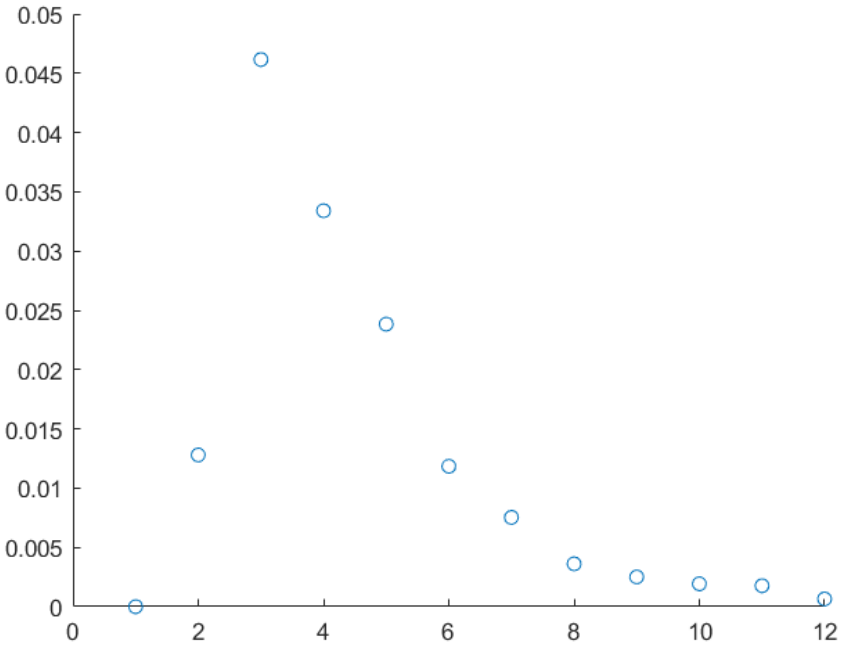


5 Counterfactual Analysis

Value of New Patient Signals Figure 5 shows the percentage differences between the adoption rates predicted by the calibrated model and when $\gamma = 1$, i.e. if new patient signals convey the same quality of information as existing patients. The quarters are displayed on the x-axis. In the first two quarters, the percentage difference in adoption rates is close to zero since new patients flow is not observed and generic prescriptions are low. As generic prescriptions start to take off and patients move in the population, the overall generic adoptions will be lower if new patient signals are seen as the same with old patient signals, with a maximum of more than 2%. This suggests that not only the volume of information matters for learning and diffusion process, but the quality of information perceived by the physicians also matters.

Information Acceleration Figure 6 shows the percentage differences in the adoption rates predicted by the calibrated model and when moving probability p is doubled. The effect of information acceleration is prominent in the early stage of the diffusion process. As information signals accumulate in the population over time, the information flow brought by new patients become less important in increasing the generic adoption rates.

Figure 6: Percentage Differences in Adoption Rates When Moving Probability Increases



6 Conclusions

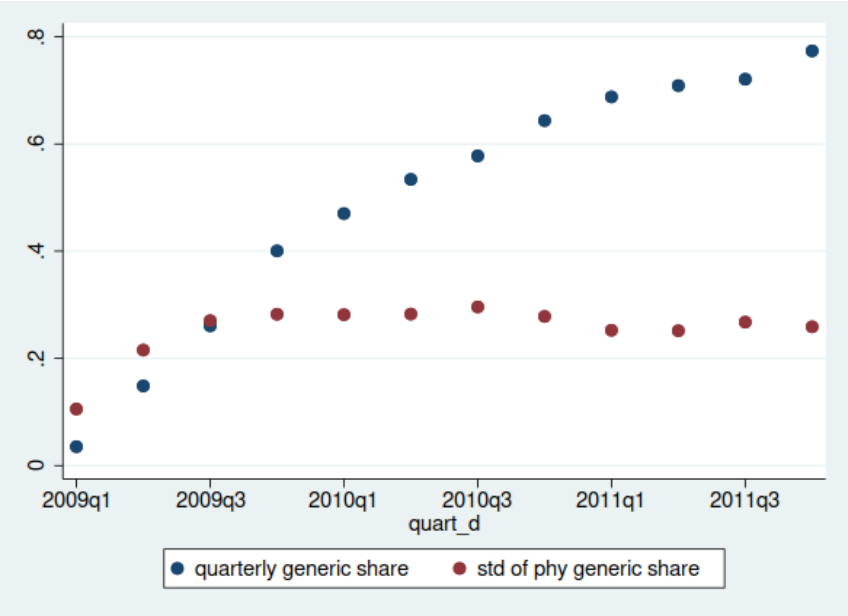
This paper examines the impacts of information on physician learning and the adoption of generic drugs. By exploiting prescription level data on the use of Atorvastatins, I find that receiving more information signals from patients on the experience of generic drugs increases the generic adoption rate of a physician. More importantly, conditional on the same volume of information, having a larger proportion of information signals from new patients who move from other physicians further increases the adoption rate of generic drugs. In addition, the variation in the generic adoption rate decreases with the number of total signals received, whilst having a higher share of new patients does not appear to affect the variations. The difference in the volume and the composition of information signals received by a physician contribute to the heterogeneity in adoption rates.

To explain the findings, I develop a Bayesian learning framework where physicians learn about generic experiences from both old existing patients and new patients from other physicians. In particular, physicians evaluate signals from old and new patients differently in the learning process. The calibrated model fits the overall diffusion pattern in the population and captures the heterogeneity in adoption rates across physicians, both in terms of individual adoption rates and the variation in adoption rates.

The counterfactual analysis highlights the role of information quality in the learning process. If the perceived informativeness does not differ across old and new patients, the diffusion will be slower in the population. In addition, if physicians are more likely

to receive information from new patients, the diffusion process can be accelerated in the population. The increase in the information flow can be due to market structure changes such as merging of healthcare centres. Policies that increase the information flow in the population can further accelerate the diffusion process, such as enabling more choices for patients.

Figure 7: Quarterly Total and S.D. of Generic Adoption Rates



A Appendix

Figure 7 shows that although generic shares increase over time, the standard deviation of physicians' generic shares within a quarter does not vary much (a slight hump shape).

Figure 8 decomposes the change in generic adoption rates across quarters:

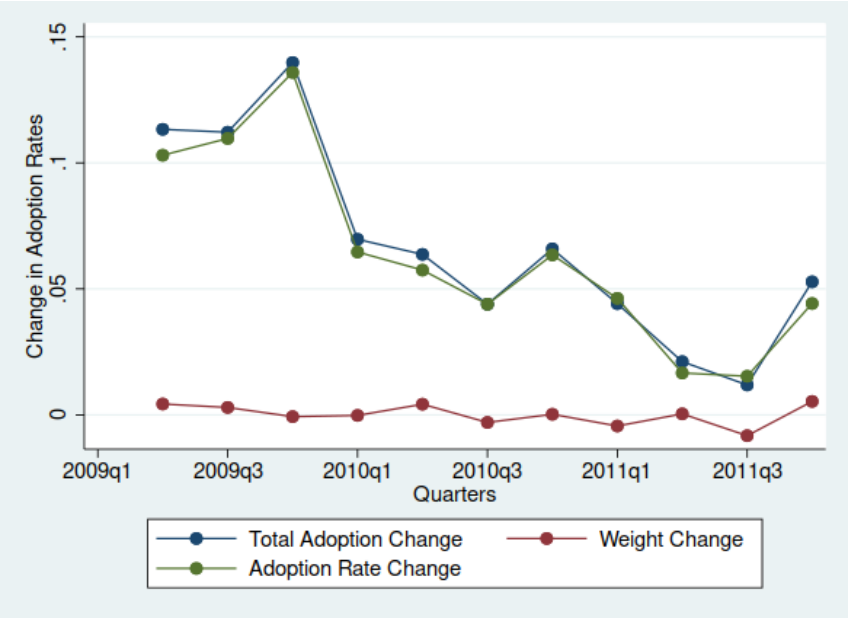
$$Adr_t - Adr_{t-1} = \sum_j w_{jt} adr_{jt} - \sum_j w_{j,t-1} adr_{j,t-1} \tag{18}$$

$$= \sum_j [\Delta w_{j,t-1} adr_{j,t-1} + \Delta adr_{j,t-1} w_{j,t-1}] \tag{19}$$

The change in generic adoptions across quarters is mainly driven by change in physicians' adoption rate, not by sorting towards physicians with high generic adoption rates.

More sets of regressions are run to validate the relationship between signal numbers and physicians' generic adoption rates.

Figure 8: Decomposition of Adoption Rate Change by Physicians



Variables	(1) $genshare_{jt}$	(2) $genshare_{jt}$
ln(total signal no.)	0.024***	0.028***
	(0.005)	(0.005)
New patient signal share		0.206***
		(0.016)
Physician Fixed Effects	Yes	Yes
Quarter Fixed Effects	Yes	Yes
Observations	14,712	14,712
R-squared	0.226	0.197

Variables	(1) $\Delta genshare_{jt}$	(2) $\Delta genshare_{jt}$
Δln(total signal no.)	0.112***	0.111***
	(0.002)	(0.002)
ΔNew patient signal share		0.228***
		(0.018)
Physician Fixed Effects	Yes	Yes
Quarter Fixed Effects	Yes	Yes
Observations	10,502	10,502
R-squared	0.312	0.323

References

Arrow, K. J., Bilir, L. K., and Sorensen, A. (2020). The Impact of Information Technology on the Diffusion of New Pharmaceuticals. *American Economic Journal: Applied Economics*, 12(3):1–39.

Bandiera, O. and Rasul, I. (2006). Social Networks and Technology Adoption in Northern Mozambique. *The Economic Journal*, 116(514):869–902.

Banerjee, A. V. (1992). A Simple Model of Herd Behavior. *The Quarterly Journal of Economics*, 107(3):797–817.

Bass, F. M. (1969). A New Product Growth for Model Consumer Durables. *Management Science*, 15(5):215–227.

Ching, A. and Ishihara, M. (2010). The Effects of Detailing on Prescribing Decisions under Quality Uncertainty. *Quantitative Marketing and Economics*, 8:123–165.

Ching, A. T. (2010). Consumer Learning and Heterogeneity: Dynamics of Demand for Prescription Drugs after Patent Expiration. *International Journal of Industrial Organization*, 28(6):619–638.

Ching, A. T., Erdem, T., and Keane, M. P. (2013). Learning Models: An Assessment of Progress, Challenges, and New Developments. *Marketing Science*, 32(6):913–938.

Ching, A. T. and Ishihara, M. (2012). Measuring the Informative and Persuasive Roles of Detailing on Prescribing Decisions. *Management Science*, 58:1374–1387.

Conley, T. G. and Udry, C. R. (2010). Learning About a New Technology: Pineapple in Ghana. *American Economic Review*, 100(1):35–69.

Coscelli, A. and Shum, M. (2004). An Empirical Model of Learning and Patient Spillovers in New Drug Entry. *Journal of Econometrics*, 122(2):213–246.

Crawford, G. S. and Shum, M. (2005). Uncertainty and Learning in Pharmaceutical Demand. *Econometrica*, 73(4):1137–1173.

DeGroot, M. H. (1974). Reaching a Consensus. *Journal of the American Statistical Association*, 69(345):118–121.

DeMarzo, P. M., Vayanos, D., and Zwiebel, J. (2003). Persuasion Bias, Social Influence, and Unidimensional Opinions. *The Quarterly Journal of Economics*, 118(3):909–968.

Ellison, G. and Fudenberg, D. (1993). Rules of Thumb for Social Learning. *Journal of Political Economy*, 101(4):612–643.

Ellison, G. and Fudenberg, D. (1995). Word-of-Mouth Communication and Social Learning. *The Quarterly Journal of Economics*, 110(1):93–125.

Epstein, A. J. and Ketcham, J. D. (2014). Information Technology and Agency in Physicians' Prescribing Decisions. *The RAND Journal of Economics*, 45(2):422–448.

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

Erdem, T., Keane, M. P., and Sun, B. (2008). A Dynamic Model of Brand Choice When Price and Advertising Signal Product Quality. *Marketing Science*, 27(6):1111–1125.

Foster, A. D. and Rosenzweig, M. R. (1995). Learning By Doing and Learning From Others: Human Capital and Technical Change in Agriculture. *Journal of Political Economy*, 103(6):1176–1209.

Granovetter, M. S. (1973). The strength of weak ties. *American journal of sociology*, 78(6):1360–1380.

Jackson, M. O. et al. (2008). *Social and economic networks*, volume 3. Princeton university press Princeton.

Mizik, N. and Jacobson, R. (2004). Are Physicians "Easy Marks"? Quantifying the Effects of Detailing and Sampling on New Prescriptions. *Management Science*, 50(12):1704–1715.

Narayanan, S. and Manchanda, P. (2009). Heterogeneous Learning and the Targeting of Marketing Communication for New Products. *Marketing Science*, 28:424–441.

Rajkumar, K., Saint-Jacques, G., Bojinov, I., Brynjolfsson, E., and Aral, S. (2022). A causal test of the strength of weak ties. *Science*, 377(6612):1304–1310.

Young, H. P. (2009). Innovation Diffusion in Heterogeneous Populations: Contagion, Social Influence, and Social Learning. *American Economic Review*, 99(5):1899–1924.