The Shifting Shape of Risk: Endogenous Market Failure for Insurance

Thomas G. Koch *

December 2008

*University of California Santa Barbara, 2127 North Hall, Santa Barbara, CA 93106; Phone: 512-809-8014; E-mail: koch.at.econ.ucsb.edu.

Abstract

Micro-foundations models typically assume that most or all of the idiosyncratic risk a household faces is uninsurable. This paper considers an economy where risk is insurable, but endogenous market failure renders some risk uninsured. I use repeated cross-sections of medical expenditures in the US to understand the role of changes in the medical risk distribution on the fraction of Americans with medical insurance. I find that both the level and the shape of the distribution of risk are important in determining the equilibrium quantity of insurance. In particular, changes to the distribution of medical risk since 1996 should have induced more risk sharing across types. Changes in the distribution associated with changing prices of medical care are primary. Changes associated with a shifting age distribution are not as important.

JEL Codes: D3, I1

Keywords: risk sharing, medical insurance, adverse selection

1 Introduction

Recent research has focused on how households insure against state-contingent losses of income by saving in non-state contingent assets, such as bonds. As in Storesletten, Telmer, and Yaron (2004), the inability to purchase state-contingent assets is an initial assumption. This restriction can be lifted, as in the case of Krueger and Perri (2006). There, markets can insure a component of the idiosyncratic risk, but are further constrained by limited enforcement of dynamic contracts. This paper considers an economy where risk is insurable, but market failure renders some risk uninsured. I follow Akerlof (1970), where adverse selection *may* lead to endogenous market failure. If the different types of risk faced are sufficiently similar, or one population is relatively large, then markets succeed, and there is risk sharing across types. While prices are not fair, they just represent transfers from the safe to the risky. If the distribution of risk types is not just so, then the market fails, and the resulting un-insurance across types reduces total welfare.

Medical expenditures represent a steadily growing portion of GDP— 16% as of 2005, and expected to grow to 20% by 2015. The price of employer-provided medical insurance to cover a single employee has grown five percent a year from 1996 to 2005. The price of family coverage has grown 5.7 percent per year over the same period.[1] At the same time, more Americans have gone without medical insurance. Among the non-poor and non-elderly, the percentage of individuals without private insurance rose 22 percent (18 percent to 22) from 1996 to 2004.[2]

One explanation for both insurance trends is the first (Swartz (2006); Freudenheim (2007)). As medical costs grow, the supply curve for medical insurance shifts up, and induces movement along the demand curve. This new equilibrium price is higher, while equilibrium quantity is lower. However, changes in the distribution of medical expenditures (e.g., mean, median, variance, etc.) are the result of changes in the distribution of medical risk. In this case, the demand curve also shifts up. The new equilibrium price is larger, but whether the new equilibrium quantity is larger or smaller is unclear.

[1]The level changes from 1996 to 2005 for these two prices is 1992 to 3227, and 4954 to 8675, as reported by the MEPS-IC. These calculations adjust for changes in the price level, as measured by the Urban CPI.

[2]These numbers are calculated from the MEPS-HC, which collects information from a nationally-representative sample of households.

This paper attempts to disentangle these changes. I use changes in the distribution of medical expenditures in the US to infer changes in the distribution of medical risk. I then take these measured differences to a quantitative model of insurance choice with endogenous price and asymmetric information. I find that measured changes in the distribution of medical risk should have made insurance more expensive and more prevalent. While the former trend is consistent with the data, the latter is not. A fraction of the falling insurance rate may be due to the tax cuts of 2001. These calculations are preceded by a series of numerical examples in a simpler setting, to provide some initial intuition for the subsequent efforts.

2 A Series of Instructive Numerical Examples

Consider an economy with two types of agents who face a Bernoulli risk—they may catch ill and have to go to the doctor. The cost of a visit, L, is the same for both types, though the healthy differ from the sick in that they face a smaller probability of going to the doctor, $p_h < p_s$. Fraction s of the population is sick, while the rest are healthy. The specific values for these parameters and calculations for each example can be found in Table 1. If preferences for risk can be characterized by CARA, with degree of risk aversion r, each agent's willingness to pay for full insurance is:

$$\pi(p_., L; r) = \frac{1}{r} \ln \left(1 - p_. + p_. e^{rL}\right)$$

Note that $\pi(p, L; r)$ is an increasing argument in each of its arguments. Sicker agents are willing to pay more for insurance, $\frac{\partial \pi(p,L;r)}{\partial p} > 0$, and all agents are willing to pay more for full insurance as the cost of a visit grows, $\frac{\partial \pi(p,L;r)}{\partial p} > 0$. It will also be useful to think of each agent's certainty equivalent, $\pi(p_., L; r) - p_. L$—the amount an agent is willing to pay for insurance above and beyond a fair price.

The average medical cost in the economy is $a^*_{all}(sp_s + (1-s)p_h)L$. Note that this value can increase because of increasing cost, L, or increasing probability of illness, $p_.$. Suppose that the population is half sick and half healthy ($s = 0.5$). Further, let $L = 1,200$, and set the probabilities of illness to $p_s = 0.5$ and $p_h = 0.1$. The average medical cost is now 360, which would be the price of insurance if all agents chose full insurance bought in competitive markets with asymmetric information—the insurers do not provide type-based insurance. However, the healthy may not choose to buy such insurance, as their willingness to pay for insurance is less than its unfair price. This is the case here, using a calibration, $r = 0.002$ from the author's previous findings—a healthy agent is only willing to pay 347 for insurance. The healthy go uninsured in such a market, while the sick sort into fairly-priced insurance.

Suppose that the price of a doctor's visit rises to $L' = 1,500$. Average medical costs rise to 450. Now note, though, that the healthy agents are willing to pay for unfairly-priced full insurance provided by a competitive market. That is, $\pi(0.1, 1200; 0.002) = 533 > 450$. As the cost of a visit grows, the certainty equivalent of the healthy agent grows faster than the price of insurance. More full insurance was provided because these changes were symmetric—both the healthy and the sick face larger risks.

Now suppose an asymmetric change, $p_s' = 0.8$, keeping the price of a doctors visit at $1,500$. Average medical costs rise to 675. A healthy agent is not willing to pay this much for full insurance, as its willingness to pay for insurance does not change. Here, there is a decrease in insurance because the change was asymmetric.

This specification for risk is clearly not useful when considering the U.S. medical insurance market, as it is very coarse and does not allow for much heterogeneity. That said, it does suggest the general principle at work in the following section: following the average medical cost is not enough to determine whether an insurance market with incomplete information will allow for more insurance or less. The change in the distribution of risk that induced the change in the average medical cost is what determines the amount of insurance provided under asymmetric information. I will now apply that general principle to the preferred specification for medical risk.

3 A More General Setting

The model economy has a unit measure of risk-averse agents with utility function $u(\cdot)$. Each agent faces medical expenditures risk, $\widetilde{mx}_i$, and its risk is private information. The risks faced by agents are heterogeneous and are distributed according to Γ. The private information assumption means that insurers are not able to sort, so all agents face a common price of insurance, ϖ. The purchase of insurance is subsidized at a rate s, which is paid for using lump-sum taxes, τ.[3] Depending upon type, each agent has a willingness to pay for insurance, π_i, which makes them indifferent between facing the risk associated with with their type, $\widetilde{mx}_i$, and paying for insurance. The decision to purchase insurance, ι_i, is a discrete choice:

$$U_I(w, \widetilde{mx}_i; \varpi, \tau) = \max_{\iota \in \{0,1\}} \left\{ u(w - (1-s)\varpi - \tau), E\Big[u(w - \tilde{mx}_i - \tau)\Big] \right\}. \tag{1}$$

If $\pi_i \geq \varpi$, then the agent purchases insurance, and does not otherwise. Because of asymmetric information, agents do not get access to fairly-priced insurance. The distribution of π_i is the demand curve for full insurance.

The market to provide full insurance is competitive, so that the price of insurance is equal to the average realized risk of the insured. Thus, I define an equilibrium to be:

- the insurance choice, ι_i, is the optimal solution to (1);
- insurance price, ϖ, is equal to the average realized risk of the insured; and,
- tax level, τ, is equal to the total cost of the subsidy of medical insurance.

The specification of preferences and risk heterogeneity is chosen to mirror several empirical facts, while also aiding computationally. The cross-sectional distribution of medical expenditures has three main features: a large number of zeros, a monotonically-decreasing partial density, and a fat, Pareto-like tail.

In order to match the three main features of the distribution of medical expenditures, I will use the following specification of individual risk and heterogeneity:

- individual i faces risk according to the exponential distribution, parameterized by λ_i; and,
- these types are distributed according to the Gamma distribution, with parameters α and β.

[3] Insurance is subsidized at a rate here as it is in the U.S. federal income tax code—here it is a positive subsidy, while it is a subsidy relative to wage income. The two are equivalent with constant absolute risk aversion.

These two assumptions lead to a cross-sectional (i.e., unconditional) Pareto-type distribution of realized medical risk, with all three characteristics. This specification of uncertainty was originally proposed by Harris (1968), though in a different context. Preferences against this risk are assumed to exhibit constant absolute risk aversion (CARA).

CARA preferences are tractable, especially in this context. The willingness to pay for insurance, given type λ_i, parameter of risk aversion r, and upper bound of risk κ can be written as:

$$\pi(\lambda_i) = \frac{1}{r} \log \left(\frac{\lambda_i - r e^{-(\lambda_i - r)\kappa}}{\lambda_i - r} \right)$$

It can be shown that this willingness to pay is monotonically decreasing in λ_i. Because of this, equilibria can be characterized by their marginal agent type, λ_m, whose willingness to pay for insurance is equal to its price. The price of medical insurance is equal to the average realized risk of the insured:

$$\begin{aligned} \varpi(\lambda_m) &= \int_0^{\lambda_m} t^{-1} \frac{t^{\alpha} e^{-\frac{t}{\beta}}}{\beta^{\alpha+1}\Gamma(\frac{\lambda_m}{\beta}, \alpha+1)} dt \\ &= \frac{\Gamma(\frac{\lambda_m}{\beta}, \alpha)}{\beta\Gamma(\frac{\lambda_m}{\beta}, \alpha+1)}, \end{aligned}$$

where $\Gamma(\cdot, \cdot)$ is the incomplete gamma function.

These specifications also mitigate the problem of multiple equilibria. Figure 1 plots out the supply and demand curves for a set of candidate parameters—the degree of risk aversion, r, upper bound of risk faced by an agent, κ, and the distribution of medical risk, α and β. Preferences and adverse selection (due to asymmetric information) lead to a downward-sloping supply curve. In this case, there are two potential equilibria. Other parameterizations could lead to one equilibrium (the supply and demand curve are just tangent), or no equilibria—the case where no risk sharing is possible.

In the face of of multiple equilibria, I introduce an equilibrium refinement.

Theorem. *An equilibrium characterized by its marginal agent* λ_m *is locally stable if for all* *local* $\varepsilon > 0$

$$\varpi(\lambda_m + \varepsilon) \geq \pi(\lambda_m + \varepsilon; \kappa)$$

and

$$\varpi(\lambda_m - \varepsilon) \leq \pi(\lambda_m - \varepsilon; \kappa)$$

The stability refinement is similar to the trembling-hand refinement of Selten (1975), and the stability refinement of Kohlberg and Mertens (1986). This refinement is intuitive—the equilibria in this model are pooling equilibria. Because the movements in insurance rates and prices are steady, it is unlikely that the equilibria we observe is unstable. Figure 1 demonstrates that the equilibrium with more insurance is the stable of the two. Parameter values that lead to a single equilibrium lead to instable equilibria.

4 Risks and Risk Sharing

The medical expenditure data used here come from the Medical Expenditure Panel Survey (MEPS). The MEPS is a series of overlapping two-year panels, with information on medical expenditures, charges, and insurance. The panel aspect of the data is beyond the focus of this paper—I use the series of cross sections in the MEPS, starting in 1996, going to 2004, in order to infer each year's distribution of medical risk—each year's Gamma distribution, its (α_t, β_t) pair. I restrict the sample to the non-elderly and the non-poor in order to avoid complications due to Medicare and Medicaid. Veterans are also removed, due to access to VA care. I will later use these values to infer the amount of insurance possible in the asymmetric-information model.

Unfortunately, the data do not present a perfect candidate for insurable medical risk. One potential measure is the total charges for an individual over a year. Because medical insurance contracts typically have co-payments or deductibles, this is an imprecise measure of the realized insurable medical risk.

This specification of risk has an internally-coherent and tractable way to back out the insurable fraction of total risk. If a fraction $\rho^{-1} < 1$ of an exponential risk is insurable, parameterized by λ, then this fraction of a risk is also an exponential risk. This new risk is parameterized by $\frac{\lambda}{\rho}$. If this fraction ρ is common across agents, then all of the exponential risks are scaled by ρ. A common ρ might come from the fact that these co-insurance rates are used to solve a hidden-action problem, and type privacy does not allow for discrimination. Finally, since β is a scale parameter for the Gamma distribution, the new distribution of insurable risk is also a Gamma distribution, with parameters α and $\frac{\beta}{\rho}$.

The distribution of insurable risk can also be inferred from the distribution of medical expenditures paid for by the privately insured. The first moment of the conditional distribution, the average realized risk of the insured, can be found by integrating the expected realized risk over the types who choose insurance; i.e.,

$$
\begin{aligned}
E(\tilde{mx}_i|\iota_i = 1) &= \int_0^{\lambda_m} t^{-1} \frac{t^{\alpha} e^{-\frac{t}{\beta}}}{\beta^{\alpha+1}\Gamma(\frac{\lambda_m}{\beta}, \alpha+1)} dt \\
&= \frac{\Gamma(\frac{\lambda_m}{\beta}, \alpha)}{\beta\Gamma(\frac{\lambda_m}{\beta}, \alpha+1)}.
\end{aligned}
$$

The average square of realized risk of the insured (i.e., the second non-central conditional moment) is found similarly:

$$
\begin{aligned}
E(\tilde{mx}_i^2|\iota_i = 1) &= \int_0^{\lambda_m} 2t^{-2} \frac{t^{\alpha} e^{-\frac{t}{\beta}}}{\beta^{\alpha+1}\Gamma(\frac{\lambda_m}{\beta}, \alpha+1)} dt \\
&= \frac{2\Gamma(\frac{\lambda_m}{\beta}, \alpha-1)}{\beta^2\Gamma(\frac{\lambda_m}{\beta}, \alpha+1)}.
\end{aligned}
$$

These are two unique moments that identify the two parameters of the risk distribution, (α, β).

That said, the amount paid for by a private insurance company is unlikely to be the amount billed to the uninsured. Insurance companies buy a large volume of medical goods and services throughout a year, and frequently bargain with with medical providers over the cost of treatment. This bargaining power is not available to the individual uninsured patient. The resolution here is the same as for the charges—assume a common proportional mark-up from the expenditures paid for by private medical insurance companies. This new marked-up distribution of risk is also a Gamma distribution, with parameters α and $\frac{\beta}{\rho}$. The balance of this paper will use both methods—marking up from private medical expenditures, and marking down from total charges. Both series tell the same story.

It may be useful to characterize changes in α and β. As the discussions of mark-up and mark-down above suggest, changes in β are easier to interpret. Since β is a scale parameter, a decrease in β scales up the medical risk faced by every agent an equal amount.

Interpreting changes in α requires a discussion of kurtosis. A distribution's kurtosis is the ratio of its fourth moment to the square of its second moment (variance). As discussed in Balanda and MacGillivray (1988), kurtosis is synonymous with "peakedness" and "tail weight." Two distributions may have the same variance, but kurtosis measures whether that variance is due to a large amount of moderate dispersion or a moderate amount of large dispersion. Flatter distributions have lower kurtosis.

Changes in kurtosis are central to the ability of insurance markets to spread risk across types. The variance of risk types may increase. If this increase is because many types became a little more different, then it is unlikely that many will newly refuse to buy the same insurance. However, if a few of the agents became very different from the rest, then cross-type risk sharing becomes less likely.

The kurtosis for a gamma distribution is $\frac{6}{\alpha}$. Thus, as α increases the risk types are becoming more and more disparate, and market failure becomes more pervasive. The sickest members of the population are becoming relatively more sick than the rest of the population. I will refer to this effect as the *kurtosis effect*—as α decreases, risk types become more similar, and insurance becomes more widespread. When κ is large, the kurtosis effect dominates.

However, the calibrated value of κ is only \$5,600. When α decreases, the marginal agent faces a larger risk, and her willingness to pay for insurance increases. However, when that agent faces only a fraction of the realized risk (i.e., κ is small), the marginal agent's willingness to pay for insurance does not increase more than the increase in the price of insurance. Here, the price of insurance is increasing because everyone who has insurance is also becoming more expensive. The kurtosis effect is dominated by the price effect of decreasing α.

Recall the claim in Swartz (2006) that increases in the average cost of care led to the increased price of insurance, and therefore fewer people had medical insurance. If the increase in average medical expenditure is due to a common increase in medical risk (β goes down), then the insurance rate may increase. As everyone's medical risk grows, so does their certainty equivalent—how much above and beyond a fair price each agent is willing to pay for insurance. Larger certainty equivalents allow for more risk sharing across types. If the increase in medical expenditure is the result of a change in the shape of the distribution of medical risk, then the outcome is ambiguous—either the price effect or the kurtosis effect may dominate.

As described in Table 2, the changes inferred from 1996 to 2004 are a mixture of the two. Over this period, we might derive these changes as coming from two sources: an increasing price of medical goods and services (falling β); and an increasing incidence of costly medical conditions (falling α). This latter trend is consistent with the aging American population and the increased incidence of obesity, each associated with costly medical conditions.

It will be assumed that there was no change in the two preference parameters over this period. The CARA risk aversion parameter, r, and upper bound of risk faced by the uninsured, κ, will be set to their values as previously estimated by the author. Saez (2004) provides the average marginal tax rates for the US. These calculations stop at the year 2000. The marginal tax rate for the year 2001 and later will be the average value for the years 1996 to 2000. Congress passed significant tax legislation in 2001, so this likely overstates the average marginal tax rate after 2001. The implications for the fit of the model are discussed below.

The mark-up rate for a given year is exactly identified as the ratio of the price of insurance and the average medical expenditure paid for insurance on behalf of the insured, both from the data. The mark-down rate from total charges for a particular year is found by matching the model's price of insurance to that found in the data. These estimates are presented in Table 2.

Figures 2(a) and 2(b) plot the insurance rate predicted by the model against the rate observed in the data. The two trends match the data well until 2001, the year of significant tax reform legislation. Since this tax reform decreased the marginal tax rates faced by households, this may explain why the model predicts increasing insurance rates, while the data finds insurance rates dropping slightly since 2000.

Figures 3(a) and 3(b) assess the relative importance of changes in the level of risk (β) and the shape of risk (α). They each plot two counterfactual insurance rates—-the first, if the shape of the distribution of risk stayed the same, but the level of risk changed as described in Table 2 (i.e., α_{1996} for all years, β_t for all t); the second, if the level of risk did not change, but the shape of the distribution of risk changed as described in the same table (i.e., β_{1996} for all years, α_t for all t).

Changes common to the medical risk of all agents are driving the insurance rate trends predicted by the model. Inferred changes in the shape in the distribution of medical risk (i.e., changes in α) play a very small roll in the predicted trends. Thus, in this period, changes due to the shifting age distribution or obesity have very little explanatory power. Instead, the model suggests a focus on changes common to all agents, such as moving prices for medical goods and services.

5 Final Remarks

An observed increase in average medical costs could be due either to a common increase in the risk faced by all; or, the shape of the distribution could have changed. A distribution's shape is best quantified by its kurtosis—is the observed variance of a distribution due to a large amount of moderate variation, or a moderate amount of large variation. In the former case, there is more opportunity for risk sharing across types. In the latter, uninsurance is more frequent, because the risk types are more disparate.

Changes in the cross-sectional distribution of medical risk suggest that there have been changes to both the level and the shape in the distribution of medical risk. The changes in the medical insurance rate predicted here are due primarily to the former. This emphasizes the role of changing medical prices for the rate of and price for insurance. It meanwhile suggests that changes due to shifting demographics, such as age and obesity, have had little effect on the number of people with medical insurance.

References

Akerlof, G. A. (1970): "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism," *The Quarterly Journal of Economics*, 84(3), 488–500.

Balanda, K. P., and H. L. MacGillivray (1988): "Kurtosis: A Critical Review," *The American Statistician*, 42(2), 111–119.

Freudenheim, M. (2007): "Cost of Health Insurance Rises Again, but at a Slightly Slower Rate," New York Times.

Harris, C. M. (1968): "The Pareto Distribution as a Queue Service Discipline," *Operations Research*, 16(2), 307–313.

Kohlberg, E., and J.-F. Mertens (1986): "On the Strategic Stability of Equilibria," *Econometrica*, 54(5), 1003–37.

Krueger, D., and F. Perri (2006): "Does Income Inequality Lead to Consumption Inequality? Evidence and Theory," *Review of Economic Studies*, 73(1), 163–193.

Saez, E. (2004): "Reported Incomes and Marginal Tax Rates, 1960-2000: Evidence and Policy Implications," NBER Working Papers 10273, National Bureau of Economic Research, Inc.

Selten, R. (1975): "Re-examination of the perfectness concept for equilibrium points in extensive games," *International Journal of Game Theory*, (4), 25–55.

Storesletten, K., C. I. Telmer, and A. Yaron (2004): "Consumption and risk sharing over the life cycle," *Journal of Monetary Economics*, 51(3), 609–633.

Swartz, K. (2006): *Reinsuring Health: Why More Middle-Class People Are Uninsured and What Government Can Do*. Russell Sage Foundation Publications.

Example	L	p_s	p_h	s	a^*_{all}	π_h	Adverse Selection?
1	1,200	0.5	0.1	0.5	360	347	Yes
2	1,500	0.5	0.1	0.5	450	533	No
3	1,500	0.8	0.1	0.5	675	533	Yes

Table 1: Parameter Values for Bernoulli Numerical Examples in the Text

Year	Charges			Expenditures			Subsidy
	α_{year}	β_{year}	ρ_{year}	α_{year}	β_{year}	ρ_{year}	
1996	1.091	5.30×10^{-4}	0.9706	1.0453	0.0010	1.8046	0.2475
1997	1.107	5.17×10^{-4}	0.968	1.1557	0.0011	2.0931	0.2533
1998	1.087	4.70×10^{-4}	0.9079	1.1645	0.0011	2.1757	0.2556
1999	1.128	4.57×10^{-4}	0.988	1.1479	0.0010	2.1994	0.2584
2000	1.156	3.95×10^{-4}	1.0028	1.1203	0.0010	2.2978	0.2613
2001	1.132	3.51×10^{-4}	0.9145	1.166	0.0008	2.118	0.2522*
2002	1.141	3.21×10^{-4}	0.9293	1.1838	0.0008	2.2149	0.2522*
2003	1.049	3.01×10^{-4}	0.84	1.0547	0.0008	1.9889	0.2522*
2004	1.067	2.72×10^{-4}	0.8126	1.1096	0.0007	2.0232	0.2522*

Table 2: Parameter values that vary by year; see text for explanation of subsidy rate for years since 2000.

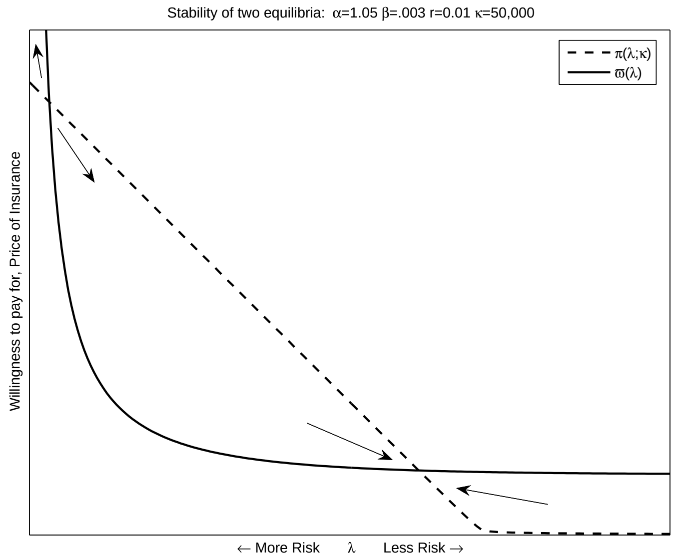
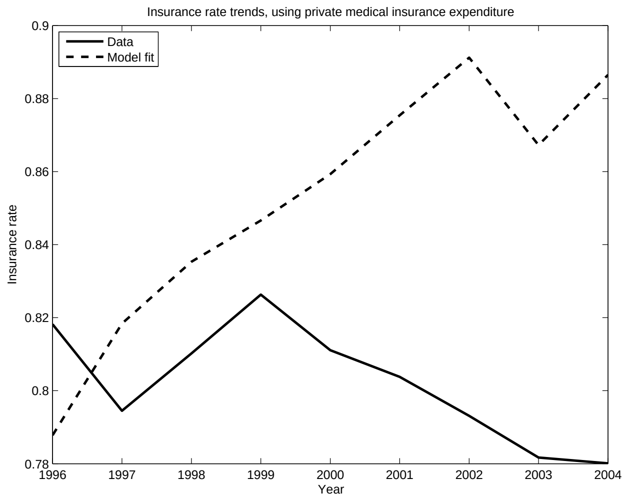
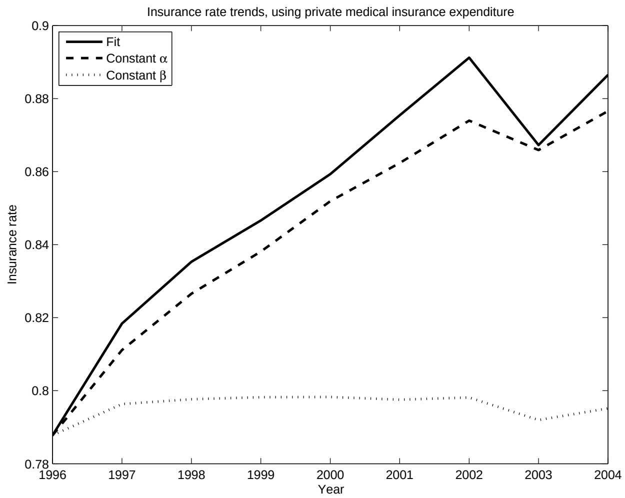


Figure 1: The stability of two equilibria.

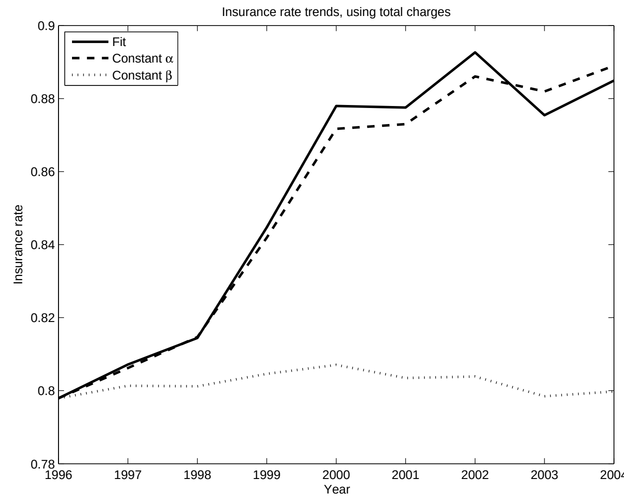


(a) (b)

Figure 2: Predicted trends are consistent across methods.



(a)



(b)

Figure 3: Changes in levels dominate changes in shape.