Moral Hazard and Dynamic Insurance Data

Jaap H. Abbring* Pierre-André Chiappori† Jean Pinquet‡

January 28, 2003

Forthcoming in the Journal of the European Economic Association

Abstract

This paper exploits dynamic features of insurance contracts in the empirical analysis of moral hazard. We first show that experience rating implies negative occurrence dependence under moral hazard: individual claim intensities decrease with the number of past claims. We then show that dynamic insurance data allow to distinguish this moral-hazard effect from dynamic selection on unobservables. We develop non-parametric tests and estimate a flexible parametric model. We find no evidence of moral hazard in French car insurance. Our analysis contributes to a recent literature based on static data that has problems distinguishing between moral hazard and selection and dealing with dynamic features of actual insurance contracts. Methodologically, this paper builds on and extends the literature on state dependence and heterogeneity in event-history data.

*Department of Economics, Free University, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands. Email: jabbring@econ.vu.nl. Jaap Abbring worked on this paper while visiting The University of Chicago, University College London and the Institute for Fiscal Studies. Abbring's research is supported by a fellowship of the Royal Netherlands Academy of Arts and Sciences.

†Department of Economics, University of Chicago, 1126 E. 59th Street, Chicago IL 60637, USA. Email: pchiappo@midway.uchicago.edu. Support from the NSF (grant # 0096516) is gratefully acknowledged.

‡Université Paris X - Nanterre.

Keywords: insurance, moral hazard, selection, occurrence dependence, event-history analysis.

JEL-codes: D82, G22, C41, C14.

We are indebted to Jim Heckman for numerous illuminating discussions. We also thank Ivar Ekeland, Bo Honoré, Bentley McLeod, Geert Ridder, Aad van der Vaart, Quang Vong, the editor (Richard Blundell), three anonymous referees, seminar participants at UCLA, Yale, Chicago, Northwestern, UCL, INSEE-CREST, Tinbergen Institute, CentER Tilburg, Toulouse, USC, Bristol and LSE, and participants in the SITE 2001 Summer Workshop and the 17th Annual Congress of the European Economic Association for helpful comments and suggestions. Errors are ours.

1 Introduction

Empirical tests of contract theory using insurance data have recently attracted much attention. Several papers test for the existence and estimate the magnitude of asymmetric-information effects in competitive insurance markets. Pueltz and Snow (1994), Dionne and Vanasse (1992), Chiappori and Salanié (1996, 2000), Dionne, Gouriéroux and Vanasse (1999, 2001) and Richaudeau (1999), to name only a few, analyze car-insurance contracts, whereas Holly et al. (1998), Chiappori, Durand and Geoffard (1998), Chiappori, Geoffard and Kyriadizou (1998), Cardon and Hendel (1998) and Hendel and Lizzeri (1999) use health or life insurance data, and Finkelstein and Poterba (2002) concentrate on annuities.

The "conditional-correlation" approach

One popular empirical strategy focuses on the correlation between the contract choice and the occurrence of an accident, conditional on observables. This correlation is informative on asymmetric-information effects. Under adverse selection, for instance, agents know whether their accident probability exceeds the average in their risk class (as defined by the insurer in terms of the available information). If it does, they are both more likely to choose a contract with more complete coverage and more likely to have an accident, everything else equal. It follows that, conditional on observables, the choice of full insurance should coincide with a higher accident rate. This is a property that can be tested using parametric or non-parametric techniques.

The conditional-correlation approach has several advantages. It is simple and very robust, as argued by Chiappori and Salanié (1996) and Chiappori et al. (2001). Furthermore, it can be used on static, cross-sectional data that are relatively easy to obtain. However, these qualities come at a cost. First, the effect of the past history of the relationship on the current contract is both difficult to model and difficult to estimate. Nevertheless, it can be of crucial importance, especially if experience rating plays a role. Second, the conditional-correlation approach may not allow to identify the type of information asymmetry involved (if any). Under adverse selection, accident-prone agents choose to buy more insurance. Moral hazard, on the other hand, suggests the opposite causality: agents who, for any reason, buy more insurance become more risky because the extensive coverage has a negative effect on incentives and discourages cautious behavior. To the extent that static data only allow to identify correlations, these two mechanisms cannot be distinguished.[1]

[1]Several articles try to empirically disentangle adverse selection and moral hazard. For instance, Holly et al. (1998) and Cardon and Hendel (1998) estimate structural models of health insurance, while Chiappori, Durand and Geoffard (1998) and Dionne et al. (2001) exploit "natural experiments" in which a new regulation exogenously changes incentives.

Adverse selection, moral hazard and the dynamics of asymmetric information

The approach used in this paper relies on the idea that adverse selection and moral hazard can be distinguished by analyzing the dynamic aspects of the relationship (Chiappori, 2000). This can be done in two different ways. One possible strategy compares the features of existing *contracts* to theoretical predictions about the form of optimal contracts under adverse selection and moral hazard. This approach exploits the fact that, in a dynamic setting, optimality has different implications in each case. Hence, a careful empirical investigation of the dynamic features of observed *contracts* may provide useful insights in the type of problem they are designed to address.[2] An alternative strategy, which we adopt throughout the present paper, does not assume optimality of existing contracts. Instead, it takes existing (and possibly suboptimal) contracts as given and contrasts the *behavior* implied by theory under adverse selection and moral hazard to observed behavior. The idea is that particular features of existing contracts, whether optimal or not, have different theoretical implications for observed behavior under adverse selection and moral hazard. Thus, the two can be distinguished by a careful analysis of observed *behavior*.

The two strategies just described have their own advantages and disadvantages. The first approach should in principle be very robust, to the extent that it relies on simple, qualitative characteristics of optimal contracts. Another virtue is that its implementation requires only data on contracts, which are in general much easier to obtain than data on induced behavior.[3] However, these qualities come at a cost. First, the qualitative characteristics of optimal dynamic contracts under asymmetric information may be very difficult to derive, except for very specific cases. They may moreover involve complex schemes, such as randomized contracts or sophisticated revelation mechanisms, which are hardly ever observed in real life.[4] A second concern is that optimal contracts are typically derived within a simplified framework (assuming for instance linear technologies, no loading, no transaction costs, etcetera). The robustness of the corresponding conclusions in a more realistic and therefore more complex setting is not guaranteed. A particularly important issue is the presence of unobserved heterogeneity in individual characteristics, notably risk aversion. Arguably, any "realistic" model of optimal insurance contracts should not only consider the particular feature under study (moral hazard or adverse selection on risk), but should also take the paramount presence of adverse selection on preferences into account. This in general requires a characterization of optimal contracts under either adverse selection and moral hazard or multidimensional adverse selection.

[2]See Dionne and Doherty (1994) for an early example.

[3]For instance, Hendel and Lizzeri (2001) find evidence of symmetric learning by comparing the actuarial value of life insurance contracts in which future premia may or may not vary with the agent's future health status. Interestingly, their analysis does *not* require data on actual mortality.

[4]For example, Chiappori et al. (1994) show that, under moral hazard, renegotiation-proof implementation of any effort level above the minimum is impossible without randomized contracts if savings are not observable. This is true except for the special case of monetary cost of effort and CARA preferences.

These are very difficult problems, especially in a dynamic context, and little is known about their solutions.[5] Finally, even casual empiricism indicates that actual insurance contracts are not always optimal (to say the least). For instance, theory suggests that the characteristics of an optimal experience-rating scheme should be specific to each class of risk; that individuals, at least in the presence of adverse selection, should be offered a menu of various experience-rating schemes; and that not only the premium, but also the deductible (and more generally the whole non-linear reimbursement profile) should depend on past experience. These features, however, are rarely observed in real life.[6]

For these reasons, we choose to adopt the second approach: we study the behavior induced by existing contracts without necessarily assuming that those contracts are optimal in any sense. Specifically, we exploit the fact that most real-life insurance contracts exhibit some form of experience rating (although not necessarily the optimal form predicted by theory). Under moral hazard, experience rating has very interesting implications. The occurrence of an accident affects the whole schedule of future premia. This changes not only the expected wealth of the agent and the expected average cost of insurance, but also, more importantly, the (expected) discounted marginal cost of future accidents. The cost of the next accident (in terms of, say, expected future premia or the corresponding certainty-equivalent) thus depends on the current premium and hence on the past accident history. It follows that the occurrence of an accident changes the incentives faced by a driver and therefore, under moral hazard, the future accident probability. This suggests that we can test for moral hazard by testing for such dynamics in the agent's accident process. A contribution of this paper is to point out the close link between this idea and a problem that has been studied at length in econometrics, the distinction between pure heterogeneity and state dependence.

Heterogeneity versus state dependence

The problem of distinguishing heterogeneity and state dependence originally appeared in economics in relation to unemployment and labor-supply issues (see Heckman and Borjas, 1980, and Heckman, 1981).[7] For example, it is well-known that individuals who are unemployed now are more likely to be unemployed in the future. There are two explanations

[5] The reader is referred to Rochet and Stole (2000) for a survey of multidimensional adverse selection, and to Chassagnon and Chiappori (1998), Jullien, Salanié and Salanié (1999) and Araujo and Moreira (2002) for examples of models involving moral hazard and adverse selection. These papers, however, only consider a static setting.

[6] From a more technical point of view, the optimality assumption also leads to difficult endogeneity issues. For instance, it is not possible, in general, to compare the performances of the different schemes that coexist on the market without taking into account the inherent selection bias: since each schedule is assumed optimal, the coexistence of different schemes must reflect differences in the corresponding populations. Such bias can be very difficult to correct for.

[7] The statistical problem of distinguishing between spurious and true state dependence has a long history, with seminal contributions by Feller (1943) and Bates and Neyman (1952).

for this empirical finding. One is that past unemployment has a direct, negative impact on the worker's future employment prospects (because of e.g. stigma effects, decreased investments in human capital, etcetera). This is the state-dependence explanation, whereby unemployment spells have a genuine effect on the agent's behavior in the sense that "an otherwise identical individual who did not experience unemployment would behave differently in the future than an individual who experienced unemployment" (Heckman and Borjas, 1980, p. 247). Alternatively, the observed pattern may simply reflect the dynamic selection effects of unobserved heterogeneity. Assume that workers differ by some unobserved characteristics that affect their employment probability. Suppose that these characteristics are positively related over time and not affected by labor-market outcomes. Then, "less employable" workers are more likely to be unemployed both now and in the future. This results in a positive association between past and future unemployment, which is spurious in the sense that past unemployment matters only as a proxy for their unobservable employability (which, by assumption, is not affected by labor-market outcomes). The literature has clarified this distinction between state dependence and pure heterogeneity and has produced various methods for their empirical analysis. In particular, panel-data methods have been developed that exploit the fact that the two explanations have different implications for the dynamics of employment.

In this paper, we will more formally develop the idea that moral hazard leads to a particular form of state dependence in the accident process under experience rating.[8] We will then argue that testing for moral hazard boils down to testing for this true state dependence in the presence of unobserved heterogeneity. We will follow the labor literature and exploit that true state dependence, and therewith moral hazard, can be detected by analyzing the dynamics of, in our case, accidents.[9]

The French "bonus-malus" scheme and state dependence

We consider a scheme used by French insurance companies, the so-called "bonus-malus" mechanism. This mechanism is both simple and explicit, which considerably simplifies the empirical investigation. Contracts are renewed and premiums are revised annually. The premium is the product of two factors, the "base premium" and the "bonus-malus coefficient" at the time of contract renewal. The base premium is computed at the beginning of the relationship. It can be defined freely, but can only depend on observables and must be uniform over agents with identical characteristics. It cannot be modified during the relationship unless some observable characteristic changes, and only in a predefined way.

[8]Note that in the contract theory framework, the theoretical structure may moreover provide specific predictions on the direction of state dependence effects. In our context, for instance, the characteristics of the experience-rating scheme at stake imply that the occurrence an accident can only increase prevention efforts.

[9]For a similar approach on labor data in a learning framework see Chiappori, Salanié and Valentin (1999).

Experience rating operates through the second component, the bonus-malus coefficient, on which we shall particularly concentrate in the paper.[10] An important feature of the French system is that the bonus-malus coefficient "sticks" to the agent, in the sense that an agent switching insurers will bring her old coefficient into the new contract. In particular, attrition cannot be explained by an attempt to "escape" the current coefficient. This will be important for the empirical analysis.

The evolution of the bonus-malus coefficient only depends on accidents for which the insuree is fully or partly responsible (accidents entirely caused by a third party are fully covered, in general by the third party's insurance and at no cost for the victim). Each year without such an accident (or, more appropriately, claim) decreases the coefficient by some fixed factor $\delta < 1$ (currently 0.95). Each accident caused by the insuree— there can be more than one in a contract year— increases the coefficient by a factor $\gamma > 1$ (currently 1.25).[11] It follows that any accident shifts the whole distribution of future (contingent) premia upwards, by a factor γ. Roughly, the "cost" of the $(n+1)$-th accident is γ times larger than that of the n-th.

These properties will, in turn, affect the optimal effort profile. A natural conjecture is that the increased marginal cost results in more cautious behavior and smaller accident probabilities. This intuition, however, deserves more careful scrutiny, because of the complex nature of the problem. Several effects should be considered. For instance, the upward shift in the premium schedule decreases the agent's expected wealth and the resulting wealth effect can modify risk aversion in a way that may confound our results. Also, the "future cost" alluded to above is in fact a random variable. Its distribution depends not only on the risk characteristics of the agent, but also on the future effort profile. Conversely, the latter will depend on (the consequences of) current behavior. In other words, the determination of the optimal effort level in each period requires the solution of an optimal control problem. In Section 2, we carefully investigate this problem by developing a theoretical model of dynamic moral hazard under experience rating. We show that, under standard convexity assumptions on preferences and the prevention technology, the intuition above is correct if the cost of insurance (premium and deductible) is a small fraction of income. Everything else equal (i.e., controlling for heterogeneity), the optimal effort level should increase with the premium. This implies that, conditionally on the driver's characteristics, the dynamics of accidents should exhibit negative "occurrence dependence": the occurrence of an accident decreases the individual's probability of future

[10] In the period covered by our data, the base premium could actually depend on the claim history as well. If anything, however, the base premium varied like the bonus-malus coefficient and amplified the experience-rating scheme.

[11] In addition, there exists a cap and a floor of the bonus-malus coefficient (currently, 3.5 and 0.5). Also, insurees that have had the maximum bonus without a claim for at least 3 years receive a "malus-deductible": their next claim at fault does not trigger a malus but only loss of the malus-deductible. It is easy to see that this does not qualitatively affect our conclusions on moral hazard and occurrence dependence.

accidents.[12] This is the true state-dependence mechanism in our insurance context. Note that this prediction relies on the presence of moral hazard.

As in the labor example above, this conclusion only holds conditionally. Unobserved heterogeneity introduces an opposite association in the raw data: good drivers, who pay lower premia, tend to have both a smaller number of past accidents and a smaller probability of future accidents. Therefore, as in the labor literature alluded to above, our main empirical task is to disentangle the effects of pure heterogeneity from those of the particular type of state dependence that is induced by the presence of moral hazard. Section 3 exploits these ideas in the empirical analysis of moral hazard using longitudinal insurance data. We specify a non-parametric econometric model of claim times that allows for both occurrence dependence and dynamic selection on unobservables. The model is specified in terms of the individual's accident (or, better, claim) rate, which is assumed to be proportional in occurrence-dependence and heterogeneity effects on the one hand and the effects of time on the other hand. Our analysis extends existing results of the state-dependence literature in various directions. We develop a model that allows for general non-stationarity (through proportional pure time effects) in the claim intensity. We propose new tests that correct for such non-stationarity, and discuss their relation to the existing literature. Finally, we analyze the identifiability of a special case of the model and present some estimation results. Section 4 concludes. Details are relegated to an appendix.

2 Dynamic moral hazard under experience rating: theory

2.1 The model

We consider a dynamic version of an insurance model along the lines of Mossin (1968). Time is continuous on $[0, \tilde{T}]$, for some finite $\tilde{T} > 0$. The wealth of agent i at time t is denoted by $W_i(t)$ and accumulates as follows. At time 0, agent i is endowed with some initial wealth $W_i(0)$. Then, between t and $t + dt$ agent i receives some income flow $w_i(t)\,dt$ and chooses a consumption flow $C_i(t)\,dt$ (where $0 \leq t < \tilde{T}$).[13] In addition, the agent causes an accident with some probability $p_i(t)\,dt$.[14] If so, she incurs some monetary loss, which is covered by an insurance contract involving a fixed deductible D_i and a premium $q_i(t)\,dt$ that is paid continuously.

The premium depends on past experience. In particular, it satisfies the following

[12] This type of state dependence is labelled "structural occurrence dependence" by Heckman and Borjas (1980).

[13] We assume that the income path is integrable on $[0, \tilde{T}]$.

[14] Accidents that are not caused by the agent are fully covered and have no impact on future premiums. Such accidents can be and are disregarded in our analysis.

"bonus-malus" system. If agent i does not cause an accident between t and $t + dt$, the premium is decreased by an amount $dq_i = \delta q_i(t)\,dt$ (the "bonus"). If she causes an accident, on the other hand, the premium jumps *discontinuously* to $\gamma q_i(t)$. Here, $\gamma - 1 > 0$ is the proportional "malus".

Accidents caused by the agent are subject to moral hazard. That is, at each time t, the agent chooses the intensity $p_i(t)$ of having an accident from some bounded interval $[0, \bar{p}_i]$, at a utility cost $\Gamma_i(p_i(t))$. We assume that Γ_i is twice differentiable on $(0, \bar{p}_i)$, with $\Gamma_i' < 0$ and $\Gamma_i'' > 0$. In words, reducing accident rates is costly and returns to prevention are decreasing. For definiteness, we also assume that $\lim_{p \uparrow \bar{p}_i} \Gamma_i'(p) = 0$.

The agent's instantaneous utility from consuming $C_i(t)$ and driving with accident intensity $p_i(t)$ at time t is $u_i(C_i(t)) - \Gamma_i(p_i(t))$. We assume that u_i is increasing and strictly concave, so that the agent is risk-averse. The agent chooses consumption and accident-intensity plans that maximize total expected utility

$$\mathbb{E}\left[\int_0^{\tilde{T}} \left(u_i(C_i(t)) - \Gamma_i(p_i(t))\right) dt\right]$$

subject to some final wealth constraint, given the wealth and premium dynamics described above. For simplicity, we assume that the agent perfectly foresees her income path $\{w_i(t); 0 \leq t \leq \tilde{T}\}$. Thus, she only has to form expectations on future accidents and their implications.

2.2 Results

For notational convenience, we now drop the index i. It should be clear, however, that all results are valid at the individual level, irrespective of the distribution of preferences and technologies across agents. In particular, the results hold for any type of unobserved heterogeneity in these primitives of the model.

A first result is

Lemma 1. *At each time t, the optimal consumption and accident intensities only depend on the past history through the agent's wealth $W(t)$ and the premium $q(t)$.*

Lemma 1 states that the only channel through which past accidents influence current behavior is their impact on the incentives faced by the agent, for which the current premium is a sufficient statistic, and on wealth. We have disregarded alternative channels, such as learning, fear, or cautionary reaction to an accident by assuming that the prevention technology, as represented by the cost function Γ, does not depend on the past experience of accidents directly. The empirical relevance of this assumption will be discussed in Section 3.

The agent faces an optimal control problem. The value function V for this problem satisfies the Bellman equation

$$\begin{aligned} V(t,W,q) = \max_{C,p} \Big\{ & u(C)\,dt - \Gamma(p)\,dt \\ & + pdtV\left(t+dt, W + w(t)dt - D - Cdt - qdt, \gamma q\right) \\ & + (1-pdt)\,V\left(t+dt, W+w(t)dt - Cdt - qdt, q - \delta qdt\right) \Big\} \end{aligned}$$

In words, between t and $t+dt$ the agent derives utility from her consumption and disutility from her prevention effort. If no accident occurs (with probability $1-pdt$), her wealth is increased by the income flow minus consumption and the premium, and the premium is continuously reduced. If the agent causes an accident (with probability pdt) then she must in addition pay the deductible, and the premium jumps to γq. The Bellman equation can be rewritten as

$$\begin{aligned} 0 = \max_{C,p} \{ & u(C) - \Gamma(p) - p\left[V(t,W,q) - V(t, W-D, \gamma q)\right] \\ & + V_t(t,W,q) + V_W(t,W,q)\left[w(t) - C - q\right] - V_q(t,W,q)\,\delta q \}, \end{aligned} \tag{B}$$

with V_t, V_W and V_q the partial derivatives of V with respect to t, W and q, respectively.

We are interested in the qualitative properties of the value function V and the optimal accident intensity. First note that agents dislike high premiums. Take some premiums q and $q' < q$. At any moment t, an agent who faces q' could derive higher utility than under the higher premium q by simply using the strategy that would be optimal under q. Using the optimal strategy at q' instead can only further improve the gain. Hence,

Lemma 2. *The value function V is decreasing in the premium q.*

We now concentrate on the impact of the current premium on the optimal accident intensity. First note that the agent's behavior for "large" premium levels may be atypical. Given the dynamics described above, the premium could exceed the agent's current income, current wealth, or even lifetime wealth if the number of accidents is large enough. This situation, however, will not arise because the agent can always choose a zero accident probability (although presumably at a high cost), say by giving up driving. Note that in that case, the accident probability becomes totally inelastic to the premium.

In most cases, however, both the premium and the deductible are "small" relative to income.[15] We now show that for such "small" values of the premium and the deductible, the prevention effort is increasing (the accident intensity is decreasing) in the premium, as intuition suggests. To this end, consider the first order conditions of the program (B),

$$\begin{aligned} u'\left(C^*(t,W,q)\right) &= V_W(t,W,q) \\ -\Gamma'\left(p^*(t,W,q)\right) &= V(t,W,q) - V(t, W-D, \gamma q), \end{aligned}$$

[15] Both premiums and deductibles are a few hundred dollars in our sample, which is one or two percent of the median household income in the population under consideration.

where $p^*(t, W, q)$ and $C^*(t, W, q)$ are, respectively, the optimal accident and consumption intensities at time t, wealth W and premium q. The first equation is the standard Euler condition for intertemporal optimality. The second condition implies that

$$\Gamma''(p^*(t, W, q))\, p_q^*(t, W, q) = \gamma V_q(t, W - D, \gamma q) - V_q(t, W, q),$$

where p_q^* is the partial derivative of p^* with respect to q. For "small" values of q and D, this condition becomes

$$p_q^*(t, W, q) \approx \frac{\gamma - 1}{\Gamma''(p^*(t, W, q))} V_q(t, W, 0) < 0$$

and the accident intensity decreases with the premium.

As a consequence, the intensity of the accident process will drop discontinuously at the time of an accident, in response to a discontinuous jump in the premium. We formally state this conclusion as

Proposition 1. *For small enough values of the premium and the deductible, the optimal accident intensity drops discontinuously at the time of an accident.*

Proposition 1 provides a simple testable implication of moral hazard under experience rating: under moral hazard, the occurrence of an accident results in a discontinuous drop in the accident intensity. In other words, the accident process should exhibit *negative occurrence dependence*. The rest of the paper is devoted to empirical tests of this prediction.

It is important to stress again that the theoretical analysis in this section operates at the individual level. For any given agent i the dynamics of accidents should exhibit the type of state dependence just described, irrespective of the distribution of preferences and technologies across agents. However, empirical tests of occurrence dependence have to rely on inter-individual comparisons, if only to control for time-effects that are common across individuals. Unobserved heterogeneity thus becomes a critical issue in the empirical analysis.

The theoretical model provides a simplified representation of actual experience-rating schemes and the agent's behavior under these schemes. We already mentioned the fact that we ignore the non-monetary consequences of an accident. A more technical issue is that the theory assumes that the premium is adjusted continuously, whereas premiums are typically only adjusted at discrete dates (e.g., annually) in actual experience-rating systems. Finally, real-life schemes often entail ceilings and floors on the premium.[16] Our view is that the previous model nevertheless provides a useful approximation of actual behavior.

[16] In France, for instance, the bonus-malus coefficient cannot fall below 0.5 or increase above 3.5. Given the actual values of γ (1.25) and δ (0.95 annually), however, the 0.5 level cannot be reached within 13 years. The 3.5 coefficient can be reached more quickly, but requires at least 6 accidents (and more if the agent receives bonuses for claim-free years before having 6 accidents). Therefore, given that drivers have on average one accident every 7 years, reaching the ceiling quickly is a very rare event. Indeed, we do not find any driver in our data who is at the ceiling.

3 Econometric specification and empirical analysis

3.1 Introduction

We now turn to the main problem of this paper, the empirical distinction of moral hazard on the one hand and adverse selection or, more generally, selection induced by unobserved heterogeneity on the other hand. As argued in the introduction, we propose to exploit dynamic data on claims under experience-rated contracts. Proposition 1 suggests a simple, direct test on negative occurrence dependence in observed claim rates. However, we can only directly control for observed individual characteristics and the occurrence-dependence effects due to moral hazard are likely to be confounded with the effects of dynamic selection on unobservable characteristics. In particular, insurees with a history of many accidents are likely to be more accident-prone for unobserved reasons. This leads to a positive effect of past claims on current claim probabilities that counters the negative effect of any moral hazard. In other words, the problem of distinguishing between moral hazard and selection is similar to a standard problem of distinguishing state (occurrence) dependence and unobserved heterogeneity.

The solution to this problem depends primarily on the type of data available. In many empirical studies in insurance (including ours), data are derived from the insurance companies' administrative files. Many relevant characteristics of the driver (age, gender, place of residence, type of job, etcetera) and the car (brand, model, vintage, power, etcetera) are used by companies for pricing purposes. These data are typically available to the econometrician as well. The same is true for the characteristics of the contract (type of coverage, premium, deductible, etcetera). Finally, each accident— or more precisely each claim– is recorded with all the relevant background information.

Data sets mainly differ in the way the past claim history of insurees is recorded and made available to researchers. Many experience-rating schemes (including the French) can be implemented with data on the number of claims in each contract year only. Often this results in only (panel) counts of claims being provided to econometricians, without any information on the exact dates of the accidents. Some schemes can even be implemented if only the total number of claims in a given period is known. This gives rise to (cross-sectional) claim-count data. The empirical distinction of state-dependence and heterogeneity with such (panel or cross-sectional) count data is a difficult problem, but one that has been studied in the literature. However, we leave application and extension of this literature to our insurance problem for future work.

In this paper, we study the more favorable situation in which the exact date of each claim is provided. This suggests specifying a continuous-time event-history model of claims that allows for occurrence dependence and unobserved heterogeneity. The theory of Section 2 shows that moral hazard leads to negative dependence of individual claim intensities on the occurrence of previous claims. Unobserved heterogeneity in claim intensities captures any dynamic selection effects. In general, the occurrence-dependence effects of moral hazard will be heterogeneous in the population. To accommodate this, we allow for general interactions between occurrence dependence and individual-specific effects in most of our analyses. We will derive some extra, stronger results for an econometric model in which occurrence dependence acts proportionally on the claim intensity and is homogeneous across agents.

Three caveats should be mentioned. First, premiums are only updated annually and not continuously as in the theoretical model. Under discounting, this may introduce non-stationarity in the agent's decision problem beyond that implied by the finite horizon in the theoretical model: when contract renewal is near, the "cost" of an accident is higher than when premiums have just been updated. This effect seems to be of only minor importance. It does however suggest that we should allow for non-stationarity in "contract time" (i.e., time since the last premium update).[17] This is particularly true because contract-time effects may bias our assessment of occurrence dependence in arbitrary ways.

Second, we observe only claims, not accidents, and the decision to file a claim is endogenous. It is well known that this introduces a second type of moral hazard, which is often referred to as "ex post" in the insurance literature. For instance, experience rating results in more cautious driving ex ante *and* increased reluctance to file a claim for a minor accident ex post.[18] Neither the theory nor the empirical analysis distinguishes between these two types of moral hazard (in a sense, the prevention technology in the theoretical model can be seen as a reduced form for both). It is not clear, actually, that ex-ante and ex-post moral hazard should be distinguished at all (typically they should not be from the insurer's perspective). Anyhow, since both effects go in the same direction, the presence of ex-post moral hazard (if any) can only bias our estimate of ex-ante moral hazard upwards. Since we fail to find any significant effect at all, we suspect that both effects are negligible.

Third, there may be learning effects. In the theory section, we assume that past accidents only affect current behavior through a monetary channel (i.e., increased monetary

[17] In the French system, claims enter the bonus-malus coefficient with a delay of 2 months. More precisely, the history considered in determining the new bonus-malus coefficient at any particular contract renewal date consists of all claims corresponding to losses incurred during the year that has ended 2 months before the renewal date. For example, a new premium that is issued on January 1, 1989 is based on the history used in writing the old (January 1, 1988) contract and all claims in the period November 1, 1987–October 31, 1988. One could say that contract time is lagging claim-history time by 2 months. Clearly, theory predicts non-stationary in claim-history time rather than in contract time. However, because the difference between the two is common across all agents, we can simply control for claim-history time by flexibly controlling for contract time.

[18] A possible solution, used by Chiappori and Salanié (2000) in their cross-sectional analysis, is to consider exclusively accidents in which a second party was involved (in which case a claim is almost automatically filled). However, this solution has two drawbacks in the present context: it would decrease the number of accidents (and especially the number of cases in which two accidents are observed) and it would lead to ignoring events (i.e., one-person accidents) that nevertheless influence incentives through their impact on the premium.

incentives to exercise caution). In reality, there may be other channels as well. In particular, a young driver is presumably not perfectly informed about her driving ability and may learn about this ability from accidents. In the presence of moral hazard, such learning may in turn affect the probability of future accidents. For example, an upward reassessment of the accident probability in the absence of effort may enhance the perceived benefits of cautious driving. We address this problem in two different ways. First, even though accidents in which the driver is at fault typically have both incentive and learning effects, accidents that are entirely caused by a third party can only have learning effects and have no monetary-incentive effects. We exploit this fact by repeating our tests on a sample including all accidents, rather than just accidents at fault. Second, learning should be more important for young drivers than for experienced drivers. Thus, we can test for the relevance of learning by repeating our analysis on sub-samples of inexperienced and experienced drivers and contrasting the results. These additional analyses, which are discussed in detail in Subsection 3.6, suggest that our conclusions are robust to learning effects.

The remainder of the paper will be concerned with the specification, identification and empirical analysis of an appropriate continuous-time model of insurance claims. We have seen that, within such a framework, a test on moral hazard boils down to a test with a null of no (genuine) occurrence dependence against the alternative of (genuine, negative) occurrence dependence. Our analysis will build on and extend the existing literature on state dependence and heterogeneity in continuous-time event-history models.

It should be stressed from the outset that the null of no moral hazard is consistent with the presence of unobserved heterogeneity, whatever its type. Such heterogeneity may reflect the impact of any information that is *not* available to the insurance company, but may or may not be known by the insuree himself. In other words, we do not, under the null, distinguish between *adverse selection* and symmetrically imperfect information.[19] It is important to note, however, that testing for adverse selection is certainly possible in this context, and can provide interesting insights in the nature of learning processes. The idea would be to analyze the changes in insurance contract initiated by the driver, and the subsequent impact on the accident hazard.[20] This is left for future work.

[19]Moreover, in most of our analyses we do not control for observed heterogeneity (see the next subsection). This observed heterogeneity will be absorbed by the individual-specific effect. Therefore, any "unobserved" heterogeneity found may also reflect symmetrically *observed* information.

[20]Chiappori and Salanié (2000) find no evidence of adverse selection on a sample of "young" (i.e., recent) drivers. They note, however, that adverse selection may also arise during the relationship, due to asymmetries in learning between the firm and the client (e.g., such an informative event as a near-miss is typically observed by the driver only). For a theoretical investigation, see de Garidel (1997).

3.2 The econometric model

Our analysis focuses on the occurrence of car insurance claims in a single insurance contract year, i.e. the period bounded by two consecutive contract renewal dates. We first present a model for the population of claim histories in the contract year.

Let time have its origin at the start of the contract year. Then, if the contract year is of length T, it can be represented by the interval $[0,T]$. Let T_k be the time of the k-th claim in the contract year. Denote the corresponding counting process by $N[0,T] := \{N(t); 0 \leq t \leq T\}$, where $N(t) := \#\{k : T_k \leq t\}$ counts the number of claims in the contract year up to time t. $N[0,T]$ is the focus of our model and empirical analysis.

The intensity θ of claims at time t, conditional on the claim history $N[0,t) := \{N(u); 0 \leq u < t\}$ up to time t and a nonnegative individual-specific effect λ is

$$\theta(t|\lambda, N[0,t)) = \lambda\, \beta(\lambda)^{N(t-)}\psi(t), \tag{M}$$

with $\beta : [0,\infty) \to (0,\infty)$ a bounded measurable function and $\psi : [0,T] \to (0,\infty)$ a continuous function that captures contract-time effects.[21] We frequently use the notation $\Psi(t) := \int_0^t \psi(u)du$. We normalize $\Psi(T) = 1$, so that λ captures the scale of θ. We assume that λ has marginal distribution G. Together with equation (M), this fully specifies the distribution of $N[0,T]$.

Equation (M) gives the *individual* claim intensity at time t of an insuree with characteristics λ and claim history $N[0,t)$. Any non-trivial dependence of this claim intensity on the insuree's claim history $N[0,t)$ can be interpreted as true state dependence. In our model (M), this state dependence takes the form of occurrence dependence: individual claim intensities at time t only depend on the claim history $N[0,t)$ through the number of past claims $N(t-)$.[22] The function β captures these occurrence-dependence effects. For an individual with characteristics λ, the claim intensity is multiplied by a factor $\beta(\lambda)$ each time a claim occurs. We allow this effect to be different across insurees with different characteristics by allowing β to be a non-trivial function of λ.

The claim intensity in (M) is multiplicative in individual heterogeneity and occurrence-dependence effects on the one hand and time-effects on the other hand. In this sense it is a proportional-hazards model. It should however be stressed that our model does not involve *observed* individual characteristics and that our main analyses do not use covariate information. Any covariate effects are subsumed in the (unobserved) individual effect λ.

[21]The model only recognizes contract time and does not explicitly consider the effects of calendar time (or duration since last event for that matter). In a typical sample, different contracts have different renewal dates, so that contract time and calendar time do not coincide (see next section). Non-stationarity in contract time arises in theory because of discounting and the discrete-time nature of contract renewal, and possibly because of learning. Also, of all time effects, contract-time effects are most likely to confound our analysis of occurrence dependence, which concerns previous occurrence of claims in the contract year. We therefore want to deal with contract-time effects in a flexible manner.

[22]Heckman and Borjas (1980) call this "structural occurrence dependence".

This individual effect can then also capture the cumulated effects of claims that have occurred before the sampled contract year started. In Subsection 3.6, we repeat our analyses on samples that are stratified on covariates. Through stratification we allow for general interactions of the covariates on the one hand and λ and the other model components on the other hand. Either way, we avoid initial-conditions problems.[23]

As explained in the Section 2, in the French car-insurance system moral hazard leads to a decline in the claim intensity with the number of previous claims: $\beta(\lambda) < 1$. Without moral hazard, we expect that $\beta(\lambda) = 1$ for all λ. Our empirical analysis focuses on distinguishing these two cases, in two stages.

First, in Subsection 3.4 we focus on testing the prediction expressed by Proposition 1 without further assumptions on preferences and technologies. In terms of the econometric model above, this amounts to testing the null hypothesis that $\beta(\lambda) = 1$ for all λ against the general moral-hazard alternative that $\beta(\lambda) < 1$ for all λ. We first state a basic result that is useful in the development of such tests: Ψ and G are identified under the null of no moral hazard. We then provide two non-parametric tests. The first of these tests, developed in Subsection 3.4.2, is rooted in the work of Bates and Neyman (1952) and Heckman and Borjas (1980, Section II.a) for exponential models with general unobserved heterogeneity. It exploits that, under the null, the total number of claims in a given (data) period is a sufficient statistic for the unobserved heterogeneity in the claim intensities. Our contribution is to provide a closely related test that allows for general non-stationarity in the claim intensities (by not imposing any restrictions on Ψ). In Subsection 3.4.3 we develop an alternative test that is inspired by the regression approach to event-history analysis (see e.g. Heckman and Borjas, 1980, Section II.b) and the methods for paired duration data developed by Holt and Prentice (1974) and Chamberlain (1985). This test uses within-individual variation in durations between claims to control for unobserved heterogeneity. The main distinguishing feature of our test is again that it allows for general non-stationarity.

Second, in Subsection 3.5 we concentrate on a particular version of the model in which the occurrence-dependence effects $\beta(\lambda)$ are the same across insurees. More formally, β is a trivial function of the individual-specific effect λ. We provide some new identification results for this model, and point out a relation to the literature on the identifiability of the two-sample mixed proportional hazard (MPH) model (Elbers and Ridder, 1982, and Kortram et al., 1995). Finally, we provide some parametric estimates of β, Ψ and G.

[23] Rather than stratifying on covariates, we could extend (M) to include proportional covariate effects. In a random-effects setting, we would typically assume that λ is independent of these covariates. However, such independence will not hold if λ captures the cumulated effects of claims that have occurred before the sampled contract year.

3.3 Sampling and data

We observe the claim histories for all insurance contracts at a French insurance company in a given and common calendar time period of two years, October 1, 1987–September 30, 1989. Different contracts have different renewal dates, so that contract time and calendar time do not coincide.

The length of the contract year is, appropriately, one year (i.e., $T = 1$, with time measured in years).[24] Ideally, contracts cannot be terminated within a contract year, except in special circumstances. An example is death of the agent. Contracts can however be changed during the year. For example, the deductible can be altered or the contract can be transferred to a new car, which may be in a different risk class. In our data set, each change of contract triggers creation of a new record. Records have to be consolidated back into single contracts. Fortunately, high quality information is available to facilitate such linking. However, some linking problems may remain and generate some spurious attrition. Overall, 9.4% of the contracts written or renewed in the first sample year fail to survive a full contract year after that.[25] In this paper, we ignore both true and spurious attrition.

With that qualification, we observe at least one full contract year for each contract that is not terminated during the first sample year or written anew during the second. For contracts with renewal dates coinciding with the start date of the sample, we observe two full contract years if the contract is renewed after one year. Our observations are the flow of contracts that are written or renewed in the first sample year, and each observation provides information on $N[0,T]$ for the contract year following the renewal or writing of the contract.[26] Consistently with the discussion above, contracts suffering from attrition during the contract year are discarded. For contracts with two contract years in the sampling period, the second year is discarded.

Let the resulting random sample contain n contracts, labelled $1, \ldots, n$. The i-th observation in the sample is denoted by $N_i[0,T]$, $i = 1, \ldots, n$. Each claim history $N_i[0,T]$ in the sample can alternatively be characterized as the number of claims $N_i(T)$ with, if $N_i(T) > 0$, a vector $(T_{1,i}, \ldots, T_{N_i(T),i})$ of claim times. Each observation $N_i(0,T)$ is complemented with an unobserved effect λ_i that has distribution G. The claim intensity for observation i is assumed to be given by (M) evaluated at sample variables.

The bottom panel of Table 1 provides some information on the sample. Accidents are fairly rare, but the number of contracts n for which we have at least one full contract year

[24]More precisely, given the presence of a leap year in the sample period, we take it to be 365 days.

[25]More precisely, we compute the rate of attrition as a fraction of all the contracts that are active at their renewal date in the first sample year, and survive for at least 31 days after that. The latter condition excludes contracts that are terminated at their renewal date in the first sample year, with a 31 day grace period.

[26]Recall from footnote 17 that claims in the last 2 months of the contract year $[0,T]$ do not affect the new premium that will be issued at T, but only the premium that may be issued one year later, at $2T$. We sample contract years to avoid attrition problems.

is large, 79,684. Of these contracts, 4,831 have one claim in the contract year, 270 have two claims, 15 have three claims, and 2 have four claims. No contracts have more than four claims.

One additional subtlety should be discussed here. The data distinguish between various types of claims. Two types of claims are labeled to be at fault, either full (inducing a 25% premium increase) or partial (12.5%), and are directly relevant to our analysis of moral hazard. We will not distinguish between these two types of claims and treat each claim, either partially or fully at fault, to be an event counted by $N(t)$. The sensitivity of the results with respect to these and other choices is investigated by repeating the analyses on other samples in Subsection 3.6.

More generally, note that we only model part of the relevant events at this point. We do not deal with changes of contract as described above, and we ignore claims that are not at fault (and, sometimes, claims at partial fault). To some extent, we could deal with these events by including them as time-varying covariates. However, because of the endogenous nature of, for example, changes in contract, we plan to pursue a more structural approach, leading to a richer event-history specification. This is left for future research.

3.4 Testing for moral hazard

3.4.1 Identification and estimation of Ψ under the null of no moral hazard

We will first show that the contract-time function Ψ can be identified and estimated under the null hypothesis that $\beta(\lambda) = 1$ for all λ (which, in the sequel, we simply denote by $\beta = 1$). This result will be useful later. As a by-product, we will be able to discuss and apply a well-known test for occurrence dependence due to Bates and Neyman (1952) and Heckman and Borjas (1980). Our first moral-hazard test, which will be discussed in next subsection, is based on this test. Appendix A.1 provides details that are omitted from this subsection.

Let H_1 be the distribution of the first claim time T_1 in the subpopulation with exactly one claim in the contract year ($N(T) = 1$):

$$H_1(t) = \Pr(T_1 \leq t | N(T) = 1).$$

Clearly, H_1 is identified from the distribution of the claim history $N[0, T]$ and can be estimated consistently by the empirical analog of H_1,

$$\hat{H}_{1,n}(t) = \frac{1}{M_{1,n}} \sum_{i=1}^{n} I\left(T_{1,i} \leq t, N_i(T) = 1\right).$$

Here, $M_{k,n} := \sum_{i=1}^{n} I\left(N_i(T) = k\right)$ more in general denotes the number of contracts in the sample with exactly k claims. It is easy to show that

$$H_1(t) = \frac{\Psi(t)}{\Psi(T)} = \Psi(t) \qquad \text{(H1)}$$

under the null hypothesis that $\beta = 1$. Note in particular that (H1) does not involve the distribution G of the individual-specific effect λ. This reflects the fact noted earlier that $N(T)$ is sufficient for λ under the null. It follows that Ψ is identified directly from H_1 and can be consistently estimated by $\hat{H}_{1,n}$ under the null.

Note that identification of G under the null easily follows. The probability $q_0(t) := \Pr(N(t) = 0)$ of observing no claims up to time t is given by

$$q_0(t) = \mathcal{L}\left(\Psi(t)\right),$$

where $\mathcal{L}(s) := \int \exp(-\lambda s) dG(\lambda)$ is the Laplace transform of G. We have just seen that Ψ is identified from H_1 under the null. This implies that $\mathcal{L}$ is identified on $[0, 1]$ from q_0 and H_1 under the null. As a consequence, $\mathcal{L}$ and G are identified under the null.[27]

So far, we have focused on the identification of Ψ from H_1 under the null hypothesis that $\beta = 1$. If we would know Ψ, we could turn the argument around and *test* the null hypothesis that $\beta = 1$ by *testing* the equality $H_1 = \Psi$. It is easy to show that

$$H_1(t|\lambda) := \Pr(T_1 \leq t|\lambda, N(T) = 1) = \begin{cases} > \Psi(t) & \text{if } \beta(\lambda) < 1 \text{ and} \\ < \Psi(t) & \text{if } \beta(\lambda) > 1 \end{cases} \qquad (\text{H}1^\dagger)$$

for all $\lambda > 0$ and $t \in (0, T)$. Thus, a test based on the difference between $\hat{H}_{1,n}$ and Ψ could be designed to have power against the alternatives of moral hazard ($\beta(\lambda) < 1$ for all λ, or simply $\beta < 1$) and, more generally, occurrence dependence ($\beta(\lambda) \neq 1$ for some λ, or just $\beta \neq 1$).

Obviously, such a test would not be feasible because we do not know Ψ. In the literature, feasible tests have been developed under the additional assumption of stationarity (Bates and Neyman, 1952). We say that the model in (M) is "stationary" if $\psi(t)$ is constant over time t ($\psi = T^{-1}$). The assumption of stationarity simply pins down $\Psi(t)$ to be t/T, so that we can test the null hypothesis of no moral hazard by testing for uniformity of H_1. In this paper, we do not want to impose stationarity to facilitate a test on moral hazard. We will however present a uniformity test of H_1 and follow Heckman and Borjas (1980) by interpreting this as a test of the joint null hypothesis of stationarity and no moral hazard.

Standard distributional test statistics can be computed from $\hat{H}_{1,n}$. We first investigate $\hat{H}_{1,n}$ graphically. The top panel of Figure 1 plots $\hat{H}_{1,n}$ and the uniform distribution function, together with some other functions that are only of later concern. The bottom

[27] This follows successively from the real analyticity and the uniqueness of the Laplace transform (see e.g. Widder, 1946). Non-parametric *estimation* of G under the null is relatively hard and will not be pursued here. The Laplace transform $\mathcal{L}$ on $[0, 1]$ can be directly estimated from $\hat{H}_{1,n}$ and the empirical analog of q_0 on $[0, T]$. However, non-parametric estimation of $\mathcal{L}$ on $(1, \infty)$ and of G would somehow involve analytic extension and deconvolution, respectively, both of which are relatively hard to implement empirically. A computationally feasible estimator can possibly be developed along the lines of the method-of-moments estimator discussed in Heckman, Robb and Walker (1990) and in papers referenced therein.

panel graphs a kernel estimate of the density of H_1, using an Epanechnikov kernel with bandwidth 0.05. All analyses are based on the data described in the bottom panel of Table 1 and include both claims at full fault and claims at partial fault. At first glance, we find that $\hat{H}_{1,n}(t) > t/T$. If we would maintain the stationarity assumption, we could take this as evidence of moral hazard ($\beta < 1$). However, we will later conclude that the deviation of $\hat{H}_{1,n}$ from a uniform distribution should be explained by non-stationarity rather than moral hazard.

We have computed Pearson's χ^2-tests and a two-sided Kolmogorov-Smirnov (KS) test. A natural grouping for the χ^2-statistics is in 365 daily intervals. With 365 intervals, we have roughly 13 observations per interval. The χ^2-statistic for uniformity is 408.4 and has 364 degrees of freedom. The asymptotic p-value is 0.054. The p-value increases to 0.484 if time is grouped in 73 intervals of 5 days. The (two-sided) KS-test is in line with the first, finer χ^2-test, with $\sup_{t\in[0,T]} |\hat{H}_{1,n}(t) - t/T| = 0.019$ and a corresponding p-value equal to 0.058. The p-value is based on the finite-sample distribution conditional on $M_{1,n}$.

We conclude that a stationary model without occurrence dependence is only (marginally) accepted at a size of 5%.

3.4.2 Comparison of the distributions of the first and second claim times

In this subsection and Subsection 3.4.3, we concentrate on testing the null hypothesis that $\beta = 1$ against moral hazard ($\beta < 1$) or occurrence dependence ($\beta \neq 1$) without further assumptions on Ψ and G.

Analogously to H_1, define H_2 to be the distribution of the second claim time T_2 in the subpopulation with exactly two claims in the contract year ($N(T) = 2$):

$$H_2(t) = \Pr(T_2 \leq t | N(T) = 2).$$

Recall that $H_1 = \Psi$ under the null that $\beta = 1$. Now, it is easy to derive that $H_2(t) = \Psi(t)^2$ and therefore that

$$H_1(t)^2 = H_2(t)$$

for all $t \in [0, T]$ under the null. The latter equality holds for all Ψ and G. So, a feasible test that allows for general non-stationarity (Ψ) and heterogeneity (G) can be based on the difference between the empirical counterparts $\hat{H}^2_{1,n}$ and $\hat{H}_{2,n}$ of H_1^2 and H_2, where

$$\hat{H}_{2,n}(t) := \frac{1}{M_{2,n}} \sum_{i=1}^{n} I\left(T_{2,i} \leq t, N_i(T) = 2\right)$$

is defined analogously to $\hat{H}_{1,n}$. To our knowledge, such tests have not been used before.

For these tests to have power, the equality $H_1^2 = H_2$ has to break down under the alternatives of moral hazard ($\beta < 1$) and occurrence dependence ($\beta \neq 1$). A formal power analysis is beyond the scope of this paper. Instead, we will show that $H_1^2 - H_2$ is different from 0 under local alternatives to the null $\beta = 1$. Recall that β is a function, giving the occurrence-dependence effect $\beta(\lambda)$ for each individual-specific effect λ. Thus, one-sided local alternatives to the null $\beta = 1$ can be expressed as $\beta = 1 + qu$ for some bounded measurable positive function u and some small $q \in \mathbb{R}$.[28] Now consider $H_1(t)^2 - H_2(t)$ as a function of β for given t. Add the argument β and write $H_1(t;\beta)^2 - H_2(t;\beta)$ to make this explicit. We can get some idea of how $H_1(t;\beta)^2 - H_2(t;\beta)$ would change if we would move from the null $\beta = 1$ to some local alternative $\beta = 1 + qu$ by computing the "directional (Gateaux) derivative" of $H_1(t;\beta)^2 - H_2(t;\beta)$ at $\beta = 1$ in the direction u. In this case, this directional derivative is simply the ordinary derivative of $H_1(t;1+qu)^2 - H_2(t;1+qu)$ with respect to q at $q = 0$. It equals (see Appendix A.2.1)

$$\frac{d}{dq}\left[H_1(t;1+qu)^2 - H_2(t;1+qu)\right]_{q=0} = \Psi(t)^2\left(1-\Psi(t)\right)\left[\frac{-\mathcal{L}_u'''(1)}{\mathcal{L}''(1)} - \frac{\mathcal{L}_u''(1)}{-\mathcal{L}'(1)}\right], \quad \text{(G1)}$$

where $\mathcal{L}_u(s) := \int u(\lambda)\exp(-\lambda s)dG(\lambda)$.

First, consider the special case that $u(\lambda) = 1$ for all λ. Then, the derivative in (G1) represents the change in $H_1(t;\beta)^2 - H_2(t;\beta)$ in response to a small homogeneous (across λ) change in β at $\beta = 1$. In this case, $\mathcal{L}_u = \mathcal{L}$ and the factor in brackets on the right-hand side of (G1) reduces to[29]

$$\frac{-\mathcal{L}'''(1)}{\mathcal{L}''(1)} - \frac{\mathcal{L}''(1)}{-\mathcal{L}'(1)} \geq 0.$$

Thus, the right-hand side of (G1) is non-negative in this case. This suggests that moral-hazard alternatives ($\beta < 1$) near $\beta = 1$ with homogeneous β correspond to $H_1^2 < H_2$. Under homogeneous $\beta > 1$, on the other hand, we should expect that $H_1^2 > H_2$.

For general directions u this result may not hold. However, if u is positive and increasing, the derivative in (G1) is generally positive. Thus, the results for the homogeneous-β case still hold if occurrence dependence is stronger ($|\beta(\lambda) - 1|$ is larger) for high-λ agents. In particular, in the case that $\beta < 1$ and decreasing we should expect that $H_1^2 < H_2$. In words, if there is moral hazard for all agents and if high-λ (high-risk) agents are more responsive to incentives, then the results for the homogenous moral-hazard case still apply.

The top panel of Figure 1 plots $\hat{H}_{1,n}^2$ and $\hat{H}_{2,n}$. Apart from some minor reversals at the tails, we find that $\hat{H}_{1,n}^2 > \hat{H}_{2,n}$. The analysis above suggests that this is evidence of $\beta > 1$. A typical one-sided test against the moral-hazard alternative $\beta < 1$ would accept the null that $\beta = 1$ at all sizes. In retrospect, the fact that $\hat{H}_{1,n}(t) > t/T$ in Subsection 3.4.1 should be read as evidence of non-stationarity rather than moral hazard. Once we correct for both heterogeneity and non-stationarity, we do not find evidence of moral hazard.

[28] This corresponds to local alternatives such that either $\beta(\lambda) < 1$ (if $q < 0$) or $\beta(\lambda) > 1$ (if $q > 0$) for all λ and includes homogeneous-β alternatives.

[29] The inequality follows from the facts that $-\mathcal{L}'$ is completely monotonic and that completely monotonic functions are log-convex (Widder, 1946). It holds strictly unless λ is degenerate or has two points of support of which one is 0, in which case it is binding. See Lemma 4 in Appendix A.2.1.

One final concern is that we may even have evidence in favor of $\beta > 1$, for which we have not put forward any economic theory. To assess the statistical significance of the discrepancy between $\hat{H}^2_{1,n}$ and $\hat{H}_{2,n}$, we compute a two-sided KS-statistic,

$$K_n = \sup_{t \in [0,T]} \left| \hat{H}_{1,n}(t)^2 - \hat{H}_{2,n}(t) \right|$$

This test is distribution-free and finite-sample p-values (conditional on $(M_{1,n}, M_{2,n})$) are easily simulated (see Appendix A.2.2). We find $K_n = 0.067$ (p-value 0.423, conditional on $(M_{1,n}, M_{2,n}) = (4828, 272)$). So, we do not reject the null that there is no occurrence dependence at any reasonable test size.

3.4.3 Direct comparison of the first and second claim durations

The test in the previous subsection is based on a comparison of first and second claim times *across* contracts. Here, we develop a test based on a more direct comparison of the first and second claim times of each contract with two claims (or more).

Main intuition: the stationary case without censoring

To develop the main intuition, first consider the stationary model, i.e. let $\psi = T^{-1}$. Then, for given λ, T_1/T and $(T_2 - T_1)/T$ are independent exponential durations with parameters λ and $\beta(\lambda)\,\lambda$, respectively. So, we can write

$$\ln(T_1) = \ln(T) + \ln(E_1) - \ln(\lambda)$$

and

$$\ln(T_2 - T_1) = \ln(T) - \ln(\beta(\lambda)) + \ln(E_2) - \ln(\lambda)$$

for some unit exponential random variables E_1 and E_2 that are mutually independent and independent of λ. It follows that

$$\ln(T_1) - \ln(T_2 - T_1) = \ln(\beta(\lambda)) + \ln(E_1) - \ln(E_2),$$

with $\ln(E_1) - \ln(E_2)$ independent of λ and symmetrically distributed around 0. This suggests that we use within-individual variation in claim durations to learn about $\ln(\beta(\lambda))$.

Suppose we have an uncensored sample $((T_{1,1}, T_{2,1}), \ldots, (T_{1,n}, T_{2,n}))$ from the joint distribution of (T_1, T_2). Then, we can estimate $\mathbb{E}[\ln(\beta(\lambda))]$ by

$$\widehat{\ln\beta}^*_n = \frac{1}{n} \sum_{i=1}^{n} \left[\ln(T_{1,i}) - \ln(T_{2,i} - T_{1,i})\right],$$

in analogy to within-estimation with standard linear panel data, and simply base a test on $\widehat{\ln\beta}^*_n$. Such a test would be a special case of the regression test for "mean" occurrence dependence proposed by Heckman and Borjas (1980, end of Section II.b).

Alternatively, note that

$$\Pr\left(T_1 \geq T_2 - T_1|\lambda\right) = \frac{\beta(\lambda)}{1+\beta(\lambda)}.$$

If $\beta(\lambda) = 1$, either duration is equally likely to be the largest. If $\beta(\lambda) < 1$, then $\Pr\left(T_1 \geq T_2 - T_1|\lambda\right) < 1/2$, as expected. This suggests that we can base a robust test of the null $\beta = 1$ against the alternative of moral hazard on the share of contracts for which the time up to the first claim is at least as large as the duration between the first and the second claims,

$$\hat{\pi}^*_n = \frac{1}{n}\sum_{i=1}^{n} I\left(T_{1,i} \geq T_{2,i} - T_{1,i}\right).$$

This approach is somewhat reminiscent of the methods for paired duration data that have been developed by Holt and Prentice (1974), Chamberlain (1985) and Ridder and Tunalı(1999).[30] In our case, it is hard to apply these methods directly, for two reasons. First, we face a censoring problem: we only observe (at least) two spells for contracts with $T_2 \leq T$. Second, we allow for non-stationarity. The standard methods can deal with duration dependence, i.e. a common dependence of the hazards of T_1 and $T_2 - T_1$ on the duration since the contract renewal date and the first claim, respectively. This generality carries over to our robust statistic above. However, as the durations T_1 and $T_2 - T_1$ are consecutive, non-stationarity can bias our comparison in arbitrary ways. We will now investigate how we can deal with these two problems.

The censoring problem under stationarity

First, focus on the censoring problem and maintain the stationarity assumption $\psi = T^{-1}$ (Appendix A.3.1 provides details). Here, we explicitly analyze the case in which we select only contracts with exactly two claims. This has some analytical advantages, notably that the distribution of $(T_1, T_2)|(\lambda, N(T) = 2)$ does not depend on λ under the null $\beta = 1$ (again, because of the sufficiency of $N(T)$ for λ under the null). This is particularly convenient when we can adapt $\widehat{\ln\beta}^*_n$ into

$$\widehat{\ln\beta}_n = \frac{1}{M_{2,n}}\sum_{i=1}^{n} \ln\left(\frac{T_{1,i}}{T_{2,i} - T_{1,i}}\right) I(N_i(T) = 2).$$

It is easy to show that $\widehat{\ln\beta}_n$ is asymptotically normal under the null $\beta = 1$, with expectation 0 and variance $\pi^2/(3np_2)$. Here, p_k is more generally the probability that a contract has k claims in the contract year. Note that, given p_2, the asymptotic standard error does not involve (properties of) the distribution G of λ and can be consistently estimated by $\pi/\sqrt{3M_{2,n}}$.

[30]See also Van den Berg (2001) for an overview.

Next, note that under the null

$$\Pr(T_1 \geq T_2 - T_1 | \lambda, N(T) = 2) = \frac{1}{2}$$

is known and independent of λ as before. Thus, it makes sense to adapt the second test $\hat{\pi}_n^*$ into

$$\hat{\pi}_n = \frac{1}{M_{2,n}} \sum_{i=1}^{n} I\left(T_{1,i} \geq T_{2,i} - T_{1,i}, N_i(T) = 2\right).$$

Under the null $\beta = 1$, $\hat{\pi}_n$ is asymptotically normal with mean 1/2 and variance $1/(4np_2)$. The variance can be estimated consistently by $1/(4M_{2,n})$. Note that

$$\Pr(T_1 \geq T_2 - T_1 | \lambda, N(T) = 2) = \begin{cases} < \frac{1}{2} & \text{if } \beta(\lambda) < 1 \text{ and} \\ > \frac{1}{2} & \text{if } \beta(\lambda) > 1 \end{cases}$$

for all $\lambda > 0$, so that we can construct the test to have power against moral hazard ($\beta < 1$) in particular.

Appendix A.3.2 discusses the alternative case in which we select all contracts with at least two claims. The analysis of the second, robust statistic $\hat{\pi}_n$ directly extends to this case. Extending $\widehat{\ln \beta}_n$ to all contracts with at least two claims is somewhat more cumbersome because, even given $p_2, p_3, \ldots$, the asymptotic distribution of this statistic depends on (properties of) the distribution G of λ under the null. We will not pursue this here.

In our data set, we observe 270 contracts with exactly two claims. We find that $\widehat{\ln \beta}_n = -0.043$. This seems consistent with moral hazard ($\beta < 1$), but the estimated asymptotic standard error of $\widehat{\ln \beta}_n$ under the null is relatively large, 0.110. Thus, the null that $\beta = 1$ is accepted at conventional test sizes. We also find that the duration up to the first claim is larger than the duration between the first and the second claims for $\hat{\pi}_n = 50.4\%$ of the 270 contracts. If we test against the alternative of moral hazard, we do not reject $\beta = 1$ at any size. The estimated standard error of $\hat{\pi}_n$ is 3.0%, so that we do not reject against the two-sided alternative of occurrence dependence at conventional test sizes either. We can also compute the equivalent of the second statistic for all 287 contracts with at least two claims. We find that the duration up to the first claim is larger than the duration between the first and the second claims for 50.2% of these contracts, with an estimated asymptotic standard error of 3.0%. This confirms our conclusion based on $\widehat{\ln \beta}_n$ and $\hat{\pi}_n$.

A general test under non-stationarity

These results depend on the stationarity assumption $\psi = T^{-1}$. Because we have found some circumstantial evidence of non-stationarity in Subsections 3.4.1 and 3.4.2, we would like to explore the consequences of non-stationarity a bit further. First, suppose we know Ψ. Then, we can deal with possible non-stationarity by working in integrated-hazard time instead of calendar time. Note that Ψ is increasing on the supports of T_1 and T_2. So, we can work with the transformed durations

$$T_1^* = \Psi(T_1) \text{ and } T_2^* = \Psi(T_2)$$

instead of T_1 and T_2 without loss of information. This is convenient, as, for given λ, T_1^* and $T_2^* - T_1^*$ are again independent exponential random variables with parameters λ and $\beta(\lambda)\ \lambda$, respectively. We can directly apply the analysis for the exponential case above, provided that we know Ψ. Then, we can construct $T_{1,i}^* = \Psi(T_{1,i})$ and $T_{2,i}^* = \Psi(T_{2,i})$ and therefore

$$\hat{\pi}_n(\Psi) := \frac{1}{M_{2,n}} \sum_{i=1}^{n} I\left(T_{1,i}^* \geq T_{2,i}^* - T_{1,i}^*, N_i(T) = 2\right).$$

This is a generalization of $\hat{\pi}_n$ to arbitrary, but still known, non-stationarity.

In our application, we do not know Ψ and $\hat{\pi}_n(\Psi)$ is not feasible. However, recall from Subsection 3.4.1 that we can estimate Ψ consistently by $\hat{H}_{1,n}$ under the null that $\beta = 1$. This suggests substituting $\hat{H}_{1,n}$ for Ψ and using $\hat{\pi}_n(\hat{H}_{1,n})$ as our test statistic. Under the null $\beta = 1$, $\hat{\pi}_n(\hat{H}_{1,n})$ is asymptotically normal with expectation $1/2$ and variance $1/(4np_2)+1/(6np_1)$. The variance can be estimated consistently as $1/(4M_{2,n})+1/(6M_{1,n})$. Appendix A.3.3 provides details.

Substitution of $\hat{H}_{1,n}$ for Ψ comes at the price of lower power. This is due to the fact that $\hat{H}_{1,n}$ is only a consistent estimator of Ψ under the null and generally captures some of the occurrence-dependence effect if $\beta \neq 1$. We can again provide some insight by analyzing the local behavior of the test at $\beta = 1$. Let

$$\pi : h \in \mathcal{D} \mapsto \Pr\left(2h(T_1) \geq h(T_2) | N(T) = 2\right),$$

with $\mathcal{D}$ the set of all distribution functions on $[0, T]$ concentrated on $(0, T]$. Then, the population-equivalent of $\hat{\pi}_n(\hat{H}_{1,n})$ can be written as $\pi(H_1)$.

First, consider the population-equivalent $\pi(\Psi)$ of the infeasible statistic $\hat{\pi}_n(\Psi)$. As before, we can investigate the behavior of $\pi(\Psi)$, as a mapping $\beta \mapsto \pi(\Psi; \beta)$ for given Ψ, locally at $\beta = 1$. The directional derivative of $\pi(\Psi; \beta)$ at $\beta = 1$ in the direction $u > 0$ is given by

$$\frac{d}{dq}\Big[\pi(\Psi; 1 + qu)\Big]_{q=0} = \frac{1}{12} \frac{-\mathcal{L}_u'''(1)}{\mathcal{L}''(1)} > 0. \tag{G2}$$

This re-establishes, now locally at $\beta = 1$, that $\pi_n(\Psi)$ can distinguish between $\beta < 1$ and $\beta > 1$ if we know Ψ.

For $\pi(H_1)$, as a mapping $\beta \mapsto \pi(H_1(\beta); \beta)$, we have instead that

$$\frac{d}{dq}\Big[\pi(H_1(1 + qu); 1 + qu)\Big]_{q=0} = \frac{1}{12}\left[\frac{-\mathcal{L}_u'''(1)}{\mathcal{L}''(1)} - \frac{\mathcal{L}_u''(1)}{-\mathcal{L}'(1)}\right]. \tag{G3}$$

The first term is again the effect in (G2) that works through the distribution of $(T_1, T_2)|$ $(N(T) = 2)$. The second term is the counteracting effect on the time-transformation H_1. Note that the derivative in (G3) has the same sign as the derivative of $H_1^2 - H_2$ in Subsection 3.4.2. It follows that generally $\pi(H_1) < 1/2$ for homogeneous moral hazard alternatives close to the null. Furthermore, this result carries over to the moral-hazard alternative in which high-risk (high-λ) agents are more responsive to incentives (that is, have higher $|\beta(\lambda) - 1|$). Appendix A.3.4 provides details.

We find that $\hat{\pi}_n(\hat{H}_{1,n}) = 50.7\%$ in our data, confirming our earlier conclusion that we do not reject the null $\beta = 1$ against the alternative of moral hazard at any test size. The estimated asymptotic standard error is 3.1%, so that we do not reject the null against a two-sided alternative at reasonable sizes either.

3.5 Identification and estimation

3.5.1 Identification

We now specialize the model in (M) by imposing homogeneity of the occurrence-dependence effects across contracts. Formally, we impose that β is a trivial function of λ and simply write

$$\theta(t|\lambda, N[0,t)) = \lambda \beta^{N(t-)} \psi(t), \qquad (\mathrm{M}^\dagger)$$

with β now a positive scalar parameter. This additional structure will facilitate the identification and estimation of the model.

In this subsection we investigate identification. Appendix B.1 provides proofs and other details. Note that the model is fully characterized by the triple $(\beta, \Psi, \mathcal{L})$: each choice of the triple $(\beta, \Psi, \mathcal{L})$ maps into exactly one distribution of $N[0,T]$.[31] We say that $(\beta, \Psi, \mathcal{L})$ is "identified" if this mapping is one-to-one. Identification of certain features of $(\beta, \Psi, \mathcal{L})$ can be defined analogously. For example, the sign of $\beta - 1$ is identified if it is uniquely determined by the distribution of $N[0,T]$.

The following result implies that the null of no moral hazard is empirically distinguishable from the alternatives of occurrence dependence and moral hazard without further assumptions.

Proposition 2. *The sign of $\beta - 1$ is identified.*

Proof. See Appendix B.1. □

In addition, we conjecture that the parameter β is point-identified without further assumptions. Define $q_k(t) := \Pr(N(t) = k)$ for $t \in [0,T]$. Suppose that $\beta < 1$, which

[31] Recall that there is a one-to-one relation between G and its Laplace transform $\mathcal{L}$.

we can tell from the data by Proposition 2 (the case $\beta > 1$ is similar). Key to the identification of β is the fact that it satisfies

$$q_1\left\{q_0^{-1}\left[(1-\beta)q_1(t)+q_0(t)\right]\right\} = \beta q_1(t) + (1-\beta^2)q_2(t) \tag{I}$$

for all $t \in [0, T]$. Because the functions q_0, q_1 and q_2 are data, this provides a continuum of non-linear restrictions on the scalar parameter β. Further analysis of this problem is beyond the scope of this paper.

Finally, we have a result on the identifiability of Ψ and $\mathcal{L}$ in the case that β is known. First note that we then also know

$$\tilde{q}(t) := (1-\beta)q_1(t) + q_0(t) = \mathcal{L}(\beta\Psi(t)).$$

We have already seen in Subsection 3.4.1 that Ψ and $\mathcal{L}$ are identified if $\beta = 1$. In the case that $\beta \neq 1$, note that q_0 and $\tilde{q}$ jointly constitute the data of the restriction of a two-sample MPH model to $[0, T]$, with "treatment effect" β, "integrated baseline hazard" Ψ and Laplace transform $\mathcal{L}$ of the "mixing distribution".[32] Thus, we can identify $\mathcal{L}$ and Ψ along the lines of standard two-sample identification proofs for the MPH model (Elbers and Ridder, 1982, and Kortram et al., 1995). These proofs rely on the additional assumption that $\mathbb{E}[\lambda] < \infty$. Thus, we have

Proposition 3. *Suppose that β is known. Then, $\mathcal{L}$ and Ψ are identified in the class of models $(\beta, \Psi, \mathcal{L})$ such that $\mathcal{L}'(0+) = \mathbb{E}[\lambda] < \infty$.*

Proof. See Appendix B.1. □

The assumption that $\mathbb{E}[\lambda] < \infty$ is not innocuous. Ridder (1990) provides extensive discussion in the context of single-spell MPH models.

3.5.2 Maximum-likelihood estimation

Finally, we have estimated parametric versions of the model ($M^\dagger$) by maximum likelihood. We have chosen a piecewise-constant specification of ψ. In particular, we partition the contract year $[0, 1]$ in 12 months and set ψ to be constant within each month:

$$\psi(t) = \sum_{j=1}^{12} \psi_j I\left(\frac{j-1}{12} \leq t < \frac{j}{12}\right),$$

with $\psi_1, \ldots, \psi_{12} \geq 0$ parameters to be estimated, up to the normalization $\Psi(1) = (1/12)\cdot \sum_{j=1}^{12} \psi_j = 1$. For the distribution G of λ, we have experimented with various discrete

[32] To be precise, suppose we observe two samples of durations. Then, q_0 and $\tilde{q}$ are the survival functions in the first and second sample in the case that the units in the first sample have hazards $\psi(t)\lambda$ conditional on λ, the units in the second sample hazards $\beta\psi(t)\lambda$ conditional on λ and λ has the same distribution with Laplace transform $\mathcal{L}$ in both samples.

distributions. We have found no evidence that G has more than 2 mass points. Therefore we present results for a model without unobserved heterogeneity, in which case we only estimate a constant, and results for a model with two points of support for λ. In the latter case, we have to estimate the support points $\lambda^a, \lambda^b > 0$, and one probability $\Pr(\lambda = \lambda^a) = 1 - \Pr(\lambda = \lambda^b)$. Appendix B.2 provides details on the construction of the likelihood.

Table 1 presents results for a specification in which the unobserved heterogeneity has 2 points of support and ψ consists of 12 monthly pieces. We find an estimate of β just below 1, with a large standard error. The point estimates are consistent with non-degenerate heterogeneity, but again the precision is low. If we estimate a model without unobserved heterogeneity, the estimate of β increases to 1.729 with a relatively small standard error of 0.091. This clearly illustrates the fact that unobserved heterogeneity causes spurious positive occurrence dependence.

In the bottom panel of Figure 1 we have plotted the estimated time effects (ψ) of Table 1. The estimates closely track the kernel estimates of the density of H_1 discussed earlier. Figure 2 plots the corresponding estimate of Ψ and its pointwise 95% confidence bounds. The uniform cumulative distribution function lies well within the latter. Indeed, we do not reject stationarity according to a Wald test. If we estimate a specification with heterogeneity that imposes stationarity, the estimate of β drops to 0.817 with an estimated standard error of 0.237. At conventional sizes, however, the estimate does not deviate significantly from 1.

All in all, the results are consistent with the non-parametric tests of the previous subsection. If we only control for heterogeneity and impose stationarity, we find a point estimate of β below 1. This mirrors the finding that generally $\tilde{H}_{1,n}(t) > t/T$, which under the assumption of stationarity can be interpreted as evidence in favor of moral hazard (see Subsection 3.4.1). However, we do not find evidence of moral hazard once we control for both heterogeneity and non-stationarity. This confirms the test results in Subsections 3.4.2 and 3.4.3. It should be noted that the precision of, in particular, the maximum likelihood estimates is low if both flexible time effects and heterogeneity are included. We return to this in Section 4.

3.6 Sensitivity analysis

3.6.1 The fault-status of claims

So far, we have pooled claims at full fault and claims at partial fault, even though the financial consequences for the insuree differ quantitatively. To check whether this matters for our results, we have recomputed the tests and estimates of the previous subsections on data of claims at full fault only. There are 4,340 contracts with 1 claim at full fault, 230 contracts with 2 such claims, 11 contracts with 3 such claims and 1 contract with 4 such claims.

The test results are close to those based on all claims at fault. The stationarity tests are slightly less marginal in not rejecting stationarity, with all p-values now above 10%. All occurrence-dependence and moral-hazard tests accept the null that $\beta = 1$, with only slighly lower precisions. The three π-statistics are still just over 50%; $\widehat{\ln \beta}_n$ has switched sign to 0.026, but remains very close to 0. The estimation results confirm these results and are in line with the estimates on the pooled-claims data.

We have also rerun the tests and estimations on a data set that includes *all* claims, rather than just claims at (full or partial) fault. Even though the moral-hazard argument applies specifically to claims at fault, other theories may imply state dependence involving occurrence of not-at-fault claims as well. For example, agents may learn from accidents, even if they were not at fault, and this may lead to negative occurrence-dependence in itself (see Subsection 3.1). Obviously, given that we have not found any evidence of occurrence dependence so far, we do not expect to find any if we include all claims either. However, we should be careful as the precision of our tests and estimates will typically increase. After all, the number of claims increases considerably: there are 12,861 contracts with 1, 1,996 contracts with 2, 307 contracts with 3, 43 contracts with 4, 7 contracts with 5, 1 contract with 6 and 1 contract with 7 claims.

We do not find much evidence of non-stationarity in the raw data. The p-values of the χ^2-tests are now 0.503 and 0.063 and the p-value of the KS-test is 0.077. The KS-test of occurrence dependence is highly insignificant. Surprisingly, $\widehat{\ln \beta}_n = 0.147$ with an estimated standard error under the null of 0.041. The more robust π-statistics are however consistent with the KS-test. Using contracts with exactly 2 claims only, we find $\hat{\pi}_n = 52.0\%$ (standard error 1.1%). If we use all contracts with at least 2 claims, we find 51.8% (1.0%). The corresponding p-values for a two-sided test are 0.081 and 0.080, respectively. This may seem slightly supportive of the result for $\widehat{\ln \beta}_n$, but recall that neither of these tests corrects for non-stationarity. The most general π-test does, and delivers $\hat{\pi}_n(\hat{H}_{1,n}) = 51.2\%$ (standard error 1.2%). The two-sided p-value is 0.327.

The parametric estimation results confirm this picture. The most appropriate specification seems to be one with two-point heterogeneity and 24 (half-monthly) time-intervals. Without regressors, the estimate of β is 1.117 with a standard error of 0.131. A Wald test with 23 degrees of freedom rejects stationarity at all reasonable sizes (p-value 0.005). As expected, overall the precision is much higher, even though we now have 24 instead of 12 time intervals.

In conclusion, we do not find evidence of occurrence dependence in data including all claims, even though the results are relatively precise. One of the reasons we have put forward for (negative) occurrence-dependence effects of claims in general is learning. Thus, this suggests that there are no such learning effects. Note though that learning is particularly relevant for young and inexperienced drivers. If so, learning implies different occurrence-dependence parameters for young and old drivers. In the next subsection, we provide some results for samples stratified in this and other ways, both for data with at-fault claims only and for data with all claims.

3.6.2 Stratification with respect to some regressors

First, consider again the data with claims at (full or partial) fault. We have stratified the data on respectively sex, age and experience (driver's licence age) and have rerun the tests on each of the sub-samples.

We have 61,564 contracts with male insurees and 26,372 contracts with female insurees. The male test results closely resemble the results for males and females pooled, with the KS-test now accepting stationarity at all reasonable sizes. The results for females based on the χ^2-tests, $\widehat{\ln\beta}_n$ and the π-tests are also in line with the overall results. The precision is, understandably, low. Unlike the overall and male KS-statistics, the female KS-statistics are significant at low sizes. The KS-statistic for stationarity has a p-value of 0.004; the KS-statistic for occurrence dependence a p-value of 0.013. We find that $\hat{H}^2_{1,n} > \hat{H}_{2,n}$. Thus, for females we have an inconsistency between the χ^2-tests, $\widehat{\ln\beta}_n$ and the π-tests on the one hand and the KS-tests on the other hand. One explanation is that the low female sample size leaves room for outliers to affect the results.

Of all contracts for which the insuree's year of birth is observed, 26,372 are born in 1951 or later ("young") and 61,564 are born in 1950 or before ("old"). Recall that our data concerns contract years starting anytime during the year following October 1, 1987. We have deliberately constructed the young drivers to be truly young (and therefore a relatively small group), because we expect that any learning effects of accidents would quickly disappear with age. The test results are very similar between both age groups and are in line with the overall results. Again, the χ^2-tests are even less significant. One difference is that the KS-test on stationarity for young insurees is now highly significant, with a p-value of 0.004. This is somewhat comparable to the results for the similarly small sample of females.

These results suggest that learning effects, leading to relatively strong negative occurrence dependence for young drivers, are not important. It seems, though, that driving experience rather than age per se would interact with learning. Obviously, we do not observe actual driving experience, but we do know the years in which the insurees' driver's licences were issued. We have divided the sample in 12,712 insurees with licences issued in 1980 or later ("inexperienced") and 75,909 insurees with licences issued in 1979 or before ("experienced"). We do not find evidence of non-stationarity, although the KS-statistic for (again, the small group of) inexperienced insurees has a p-value as low as 0.082. The KS-tests on occurrence dependence are highly insignificant for either experience level, but the other occurrence-dependence tests produce some interesting results. For inexperienced drivers, we find that $\widehat{\ln\beta}_n = -0.296$ with a standard error of 0.222. The three π-statistics are in the range 40.3–41.2%, with standard errors 6.1–6.3%. For experienced drivers, on the other hand, we have that $\widehat{\ln\beta}_n = 0.041$ (0.127) and π-statistics in the range 53.0–53.7% (3.4–3.6%). The results for experienced drivers are consistent with the results for the pooled sample. Also, even if we use one-sided tests on moral hazard, we do not reject the null that $\beta = 1$ at a 5% size for inexperienced drivers. However, the differences between both experience levels are remarkable and suggest that, if anything, there is negative occurrence dependence for inexperienced drivers only. This points at learning rather than moral hazard effects of accidents.

Clearly, this conclusion cannot be drawn with any reasonable statistical significance because of the imprecision of our results. However, we have earlier argued that it may make sense to include claims that are not at fault in an analysis of learning. If this is correct, the resulting larger sample may be more informative on any differences in occurrence-dependence effects between experience levels.

The results for experienced drivers are consistent with the results for the pooled data. The occurrence-dependence tests that do not correct for non-stationarity are strongly in favor of $\beta > 1$. However, the KS-statistic on occurrence dependence is very insignificant and $\hat{\pi}_n(\hat{H}_{1,n}) = 51.8\%$ with a standard error of 1.3% and a two-sided p-value of 0.165. The results for inexperienced drivers are all insignificant, but indeed mostly pointing at $\beta < 1$. However, $\widehat{\ln \beta}_n$ is slightly positive and $\hat{\pi}_n(\hat{H}_{1,n}) = 49.6\%$ (standard error 2.8%). We conclude that the data including all claims are not supporting differences in occurrence-dependence effects between experience levels. For now, we can shelve the learning explanation of occurrence dependence.

4 Conclusion

In this paper, we show that the experience-rating structure commonly found in insurance contracts can be exploited in the empirical analysis of moral hazard. In particular, experience rating implies negative occurrence dependence of individual claim intensities under moral hazard. In other words, under moral hazard and experience rating individual claim intensities decrease with the number of past claims. In observed claim intensities, this negative occurrence dependence effect is confounded with a positive selection effect: an insuree with a large number of past claims is likely to be a bad driver and therefore to have a high future claim intensity. Thus, from an empirical perspective the distinction between moral hazard and (adverse) selection boils down to disentangling "true" state dependence and unobserved heterogeneity. This is a problem that has been studied at length in labor economics in the late 1970s and early 1980s. Following up on this literature, we develop general tests of the null of no moral hazard. Our tests are non-parametric (except for a separability assumption), and generalize existing work by allowing for general non-stationarity of the claim intensity.

We have applied our tests to French car-insurance data and have found no evidence of moral hazard (or more general occurrence dependence). More precisely, we have not rejected the null of no moral hazard against the alternatives of moral hazard or general occurrence dependence at conventional levels. This result is confirmed by parametric estimates of a flexible model that allows for both occurrence dependence and selection on unobservables.

One remaining, practical concern is that observations of multiple claims by a single insuree are central to identifying moral hazard effects. Because claims are relatively rare and we focus on claims at fault within a single contract year, we observe multiple claims for relatively few of the many contracts in our data set. This translates into a fairly low precision of our empirical results. One solution would be to resort to low-dimensional parametric models, but this would artificially generate precision at the expense of robustness. We prefer to simply qualify our identification results by noting that even a large data set carries limited information on moral hazard effects. Note that this problem does not seem to be fundamental (i.e., it would be resolved if we would have a very large data set): a complementary analysis of a larger data set of all (at-fault and not-at-fault) claims yields results of satisfactory precision.

An obvious way to expand the data is to include claim histories beyond a single contract year. Our French data provide information for up to two years and, at least in principle, it should be possible to collect alternative data on many more years. Obviously, multiple claims will be more prevalent in longer claim histories. On the downside, however, using claim histories that extend beyond a single contract year introduces the problem of dynamic contract selection. This may imply non-ignorable attrition.

This takes us to our final remark. The main contribution of this paper is to provide, in a dynamic context, tests of moral hazard that are valid in the presence of general unobserved heterogeneity. Our analysis is consistent with heterogeneity in accident rates that is symmetrically observed between the insurer and the agents. It also allows for adverse selection in the technical sense, which arises if agents are better informed about their risk than insurers. However, because we focus on claims and ignore other insurance events, we are not able to distinguish between both types of heterogeneity. In particular, we have not modelled changes in contracts (risk class, coverage, etcetera). Such changes are observed across contract years, but also within contract years. Some contract changes may be forced by events that can safely be considered to be external to the claims process, but occasionally a case can be made for the endogeneity of contract changes to the agent's claim history. A common assumption in insurance theory is that both the agent and the insurer learn about the agent's ability, but that the learning process is asymmetric because the information available to the agent is much richer. If so, one would expect the agent's decisions about contract changes to be informative about her risk, even after controlling for the information available to the insurer (i.e., the agent's observable characteristics and past history). Again, this (complex) problem is left for future research.

Appendix

A Testing for moral hazard

This appendix provides results for Subsection 3.4. Note that the analysis in this subsection is based on the general econometric model (M).

A.1 Results for Subsection 3.4.1 (H_1)

A.1.1 Asymptotic properties of $\hat{H}_{1,n}$

We use "$\Longrightarrow$" to denote convergence in distribution (weak convergence) and "$\xrightarrow{\text{a.s.}}$" to denote almost-sure convergence. Throughout, $\mathbb{G}_U$ is a uniform (on $[0,1]$) Brownian bridge and, for any distribution H, $\mathbb{G}_H = \mathbb{G}_U \circ H$ a H–Brownian-bridge. The following properties of the estimator $\hat{H}_{1,n}$ are standard.

Lemma 3.

$$\sup\left|\hat{H}_{1,n} - H_1\right| \xrightarrow{\text{a.s.}} 0 \quad \textit{and} \quad \sqrt{n}\left(\hat{H}_{1,n} - H_1\right) \Longrightarrow \frac{1}{\sqrt{p_1}}\mathbb{G}_{H_1}$$

as $n \to \infty$.

Proof. The result follows from the Glivenko-Cantelli and Donsker theorems (e.g., Van der Vaart, 1998, Theorems 19.1 and 19.3), the law of large numbers and Slutsky's lemma. □

Note that $H_1 = \Psi$ under the null that $\beta = 1$.

A.1.2 The behavior of H_1 under the alternative that $\beta \neq 1$

Under the null that $\beta = 1$, $H_1(t|\lambda) = \Psi(t)$ for $t \in [0,T]$, so that $H_1(\Psi^{-1}(z)|\lambda) = z$ for $z \in [0,1]$. If $\beta(\lambda) \neq 1$, on the other hand,

$$\Pr\left(N(t) = 1, N(T) = 1|\lambda\right) = \int_0^t \lambda\psi(u)\mathrm{e}^{-\lambda[1-\beta(\lambda)]\Psi(u)-\lambda\beta(\lambda)}du = \frac{\mathrm{e}^{-\lambda\beta(\lambda)}}{1-\beta(\lambda)}\left[1 - \mathrm{e}^{-\lambda[1-\beta(\lambda)]\Psi(t)}\right],$$

so that

$$H_1(t|\lambda) = \frac{\Pr\left(N(t) = 1, N(T) = 1|\lambda\right)}{\Pr\left(N(T) = 1|\lambda\right)} = \frac{1 - \mathrm{e}^{-\lambda[1-\beta(\lambda)]\Psi(t)}}{1 - \mathrm{e}^{-\lambda[1-\beta(\lambda)]}}.$$

Substituting $z = \Psi(t)$ gives

$$H_1(\Psi^{-1}(z)|\lambda) = \frac{1 - \mathrm{e}^{-\lambda[1-\beta(\lambda)]z}}{1 - \mathrm{e}^{-\lambda[1-\beta(\lambda)]}}.$$

$H_1(\Psi^{-1}(z)|\lambda)$ increases from 0 to 1 on $[0,1]$ and is strictly concave if $\beta(\lambda) < 1$ and strictly convex if $\beta(\lambda) > 1$. This implies that

$$H_1(\Psi^{-1}(z)|\lambda) = \begin{cases} > z & \text{if } \beta(\lambda) < 1 \text{ and} \\ < z & \text{if } \beta(\lambda) > 1 \end{cases}$$

for all $\lambda > 0$ and $z \in (0,1)$. The inequalities in (H1†) in Subsection 3.4.1 follow.

A.1.3 Kernel estimation of the density of H_1

Here, we provide the details of the estimation procedure used to estimate the Lebesgue density h_1 corresponding to H_1. A standard kernel density estimator of h_1 is

$$\tilde{h}_1(t) := \frac{1}{b}\int k\left(\frac{t-x}{b}\right) d\hat{H}_{1,n}(x) = \frac{1}{bM_{1,n}}\sum_{i=1}^{n} I\left(N_i(T)=1\right) k\left(\frac{t-T_{1,i}}{b}\right),$$

with $0 < b < 1/2$ the bandwidth and k the Epanechnikov kernel function

$$k(x) = \begin{cases} \frac{3}{4}\left(1-x^2\right) & \text{if } |x| \leq 1, \text{ and} \\ 0 & \text{if } |x| > 1. \end{cases}$$

Now, as H_1 has support $[0,T]$, $\tilde{h}_1(t)$ may have support on $[-b, T+b]$. The restriction of $\tilde{h}_1(t)$ to $[0,T]$ generally under-estimates h_1 on $[0,b]$ and $[T-b,T]$. The ad hoc solution we have used here is to "reflect" the mass of $\tilde{h}_1$ outside $[0,T]$ into $[0,T]$, and estimate h_1 by

$$\hat{h}_1(t) = \begin{cases} \tilde{h}_1(t) + \tilde{h}_1(-t) & \text{if } 0 < t < b, \\ \tilde{h}_1(t) & \text{if } b < t < 1-b, \\ \tilde{h}_1(t) + \tilde{h}_1(2T-t) & \text{if } T-b < t < T, \\ 0 & \text{if } t \leq 0 \text{ or } t \geq T. \end{cases}$$

A.1.4 Computing the distributions of the uniformity tests under the null

Time is grouped in 365 days in our sample. If we maintain that $\hat{H}_{1,n}$ is defined on the underlying continuous-time sample, this translates into observing $\hat{H}_{1,n}$ on a grid $\{T/365, 2T/365, \ldots, T\}$ only. Chi-square statistics can be straightforwardly computed using the natural grouping of the data in 365 days (or any coarser grouping). The computation of KS-statistics requires slightly more care. Grouped data only allow us to compute bounds on the continuous-time KS-statistic

$$\sup_{t\in[0,T]} \left|\hat{H}_{1,n}(t) - t/T\right|. \tag{1}$$

A sharp lower bound is given by the discrete-time KS-statistic

$$\max_{t\in\{T/365, 2T/365, \ldots, T\}} \left|\hat{H}_{1,n}(t) - t/T\right|. \tag{2}$$

An upper bound is also easy to derive and the bounds can be expected to be narrow due to the small size of the intervals relative to the density of the claims (see the similar analysis in Appendix A.2.2). We could use standard distribution theory for the continuous-time statistic in (1) to derive corresponding bounds on the p-value for this statistic. In this case, however, it is easier to simply use the discrete-time statistic in (2) itself. Its distribution under the null is known and exact critical and p-values are easy to simulate by Monte Carlo methods.

The finite-sample p-values reported are conditional on the sub-sample size $M_{1,n}$, which is random even if n is not. It is easy to see that $\hat{H}_{1,n} \sim \hat{H}^*_{1,m}$ given $M_{1,n} = m \in \mathbb{N}$, with $\hat{H}^*_{1,m}$ the empirical distribution of a random sample of fixed size m from H_1.[33] Here and below, "$\sim$" denotes equality in distribution.

A.2 Results for Subsection 3.4.2 ($H_1^2 - H_2$)

A.2.1 The behavior of $H_1^2 - H_2$ under local alternatives to $\beta = 1$

This appendix provides details on the directional derivative of $H_1(t;\beta)^2 - H_2(t;\beta)$ at $\beta = 1$ in the direction u. From Appendix A.1.2 we know that

$$H_1(t; 1+qu) = \frac{\int \frac{\mathrm{e}^{-\lambda[1+qu(\lambda)]}}{qu(\lambda)} \left[\mathrm{e}^{\lambda qu(\lambda)\Psi(t)} - 1\right] dG(\lambda)}{\int \frac{\mathrm{e}^{-\lambda[1+qu(\lambda)]}}{qu(\lambda)} \left[\mathrm{e}^{\lambda qu(\lambda)} - 1\right] dG(\lambda)}.$$

It is easy to derive that

$$\frac{d}{dq}\Big[\int \frac{\mathrm{e}^{-\lambda[1+qu(\lambda)]}}{qu(\lambda)} \left(\mathrm{e}^{\lambda qu(\lambda)\Psi(t)} - 1\right) dG(\lambda)\Big]_{q=0} = -\frac{1}{2}\Psi(t)\left[2 - \Psi(t)\right] \mathcal{L}_u''(1),$$

so that

$$\frac{d}{dq}\Big[H_1(t;1+qu)^2\Big]_{q=0} = -\Psi(t)^2\left[1 - \Psi(t)\right] \frac{\mathcal{L}_u''(1)}{-\mathcal{L}'(1)}.$$

Next, for λ such that $\beta(\lambda) \neq 1$, we have that

$$\begin{aligned}
&\Pr\left(N(t) = 2, N(T) = 2|\lambda\right)\\
&= \int_0^t \int_{t_1}^t \lambda^2 \beta(\lambda)\psi(t_1)\psi(t_2)\mathrm{e}^{-\lambda[1-\beta(\lambda)]\Psi(t_1) - \lambda\beta(\lambda)[1-\beta(\lambda)]\Psi(t_2) - \lambda\beta(\lambda)^2} dt_2 dt_1\\
&= \frac{\mathrm{e}^{-\lambda\beta(\lambda)^2}\left(1 - [1+\beta(\lambda)]\mathrm{e}^{-\lambda\beta(\lambda)[1-\beta(\lambda)]\Psi(t)} + \beta(\lambda)\mathrm{e}^{-\lambda[1-\beta(\lambda)^2]\Psi(t)}\right)}{[1-\beta(\lambda)]^2[1+\beta(\lambda)]}.
\end{aligned}$$

---

[33] For $l \in \mathbb{N}$ and $(t_1, \ldots, t_l) \in \mathbb{R}^l$, we can write

$$\begin{aligned}
&\Pr\left(\hat{H}_{1,n}(t_1) \leq x_1, \ldots, \hat{H}_{1,n}(t_l) \leq x_l \,\Big|\, M_{1,n} = m\right)\\
&= \mathbb{E}\left[\Pr\left(\hat{H}_{1,n}(t_1) \leq x_1, \ldots, \hat{H}_{1,n}(t_l) \leq x_l \,\Big|\, N_1(T), \ldots, N_n(T), M_{1,n} = m\right) | M_{1,n} = m\right].
\end{aligned}$$

Note that this holds in particular for $l = 365$ and $t_j = Tj/365$. Because

$$\begin{aligned}
&\Pr\left(\hat{H}_{1,n}(t_1) \leq x_1, \ldots, \hat{H}_{1,n}(t_l) \leq x_l \,\Big|\, N_1(T) = n_1, \ldots, N_n(T) = n_n\right)\\
&= \Pr\left(\frac{\sum_{i=1}^m I(T_{1,i} \leq t_1)}{m} \leq x_1, \ldots, \frac{\sum_{i=1}^m I(T_{1,i} \leq t_l)}{m} \leq x_l \,\Bigg|\, N_1(T) = \cdots = N_m(T) = 1\right)
\end{aligned}$$

for each $(n_1, \ldots, n_n) \in \{0,1\}^n$ such that $\sum_{i=1}^n n_i = m$, it follows that

$$\begin{aligned}
&\Pr\left(\hat{H}_{1,n}(t_1) \leq x_1, \ldots, \hat{H}_{1,n}(t_l) \leq x_l \,\Big|\, M_{1,n} = m\right)\\
&= \Pr\left(\frac{\sum_{i=1}^m I(T_{1,i} \leq t_1)}{m} \leq x_1, \ldots, \frac{\sum_{i=1}^m I(T_{1,i} \leq t_l)}{m} \leq x_l \,\Bigg|\, N_1(T) = \cdots = N_m(T) = 1\right).
\end{aligned}$$

Thus, for $q \neq 0$

$$H_2(t; 1+qu) = \frac{\int \frac{e^{-\lambda[1+qu(\lambda)]^2}}{q^2u(\lambda)^2[2+qu(\lambda)]}\left(1-[2+qu(\lambda)]e^{\lambda qu(\lambda)[1+qu(\lambda)]\Psi(t)}+[1+qu(\lambda)]e^{\lambda qu(\lambda)[2+qu(\lambda)]\Psi(t)}\right)dG(\lambda)}{\int \frac{e^{-\lambda[1+qu(\lambda)]^2}}{q^2u(\lambda)^2[2+qu(\lambda)]}\left(1-[2+qu(\lambda)]e^{\lambda qu(\lambda)[1+qu(\lambda)]}+[1+qu(\lambda)]e^{\lambda qu(\lambda)[2+qu(\lambda)]}\right)dG(\lambda)}.$$

Now, using that

$$\frac{d}{dq}\left[\int \frac{e^{-\lambda[1+qu(\lambda)]^2}}{q^2u(\lambda)^2[2+qu(\lambda)]}\left(1-[2+qu(\lambda)]e^{\lambda qu(\lambda)[1+qu(\lambda)]\Psi(t)}+[1+qu(\lambda)]e^{\lambda qu(\lambda)[2+qu(\lambda)]\Psi(t)}\right)dG(\lambda)\right]_{q=0} = \frac{1}{2}\Psi(t)^2\left\{[2-\Psi(t)]\,\mathcal{L}_u'''(1)+\mathcal{L}_u''(1)\right\}$$

we find that

$$\frac{d}{dq}\Big[H_2(t;1+qu)\Big]_{q=0} = -\Psi(t)^2\,[1-\Psi(t)]\,\frac{-\mathcal{L}_u'''(1)}{\mathcal{L}''(1)}.$$

Conclude that

$$\frac{d}{dq}\Big[H_1(t;1+qu)^2 - H_2(t;1+qu)^2\Big]_{q=0} = \Psi(t)^2\,(1-\Psi(t))\left[\frac{-\mathcal{L}_u'''(1)}{\mathcal{L}''(1)} - \frac{\mathcal{L}_u''(1)}{-\mathcal{L}'(1)}\right]. \tag{G1}$$

The following two results provide sufficient conditions for (G1) to be positive on $(0, T)$. First, if u is constant (trivial) then $\mathcal{L}_u^{(k)}(1) = u\mathcal{L}^{(k)}(1)$ (here, the superscript (k) denotes the k-th derivative). Then, Lemma 4 implies that (G1) is positive if u is positive and G has at least two positive points of support.

Lemma 4. *If $\mathcal{L}$ is the Laplace transform of a distribution G with nonnegative support such that $G(0) < 1$, then*

$$-\frac{d}{ds}\ln\left[\frac{\mathcal{L}''(s)}{-\mathcal{L}'(s)}\right] = \frac{-\mathcal{L}'''(s)}{\mathcal{L}''(s)} - \frac{\mathcal{L}''(s)}{-\mathcal{L}'(s)} \geq 0, \tag{3}$$

with equality holding if and only if G is either degenerate or has two points of support of which one is 0.

Proof. For given $s \in (0,\infty)$, $x \in [0,\infty) \mapsto \mathcal{L}(s+x)/\mathcal{L}(s)$ is the Laplace transform of a distribution $\tilde{G}$ that has the same support as G. The k-th moment of $\tilde{G}$ exists and is given by $\tilde{\mu}_k := (-1)^k\mathcal{L}^{(k)}(s)/\mathcal{L}(s)$. Thus, (3) has the sign of $\tilde{\mu}_3\tilde{\mu}_1 - (\tilde{\mu}_2)^2$ and the claimed result follows from standard results for the Stieltjes moment problem (e.g. Shohat and Tamarkin, 1943, Theorem 1.3). □

Second, for positive and increasing u we can apply

Lemma 5. *If u is positive and increasing, then*

$$\frac{-\mathcal{L}_u'''(1)}{\mathcal{L}''(1)} - \frac{\mathcal{L}_u''(1)}{-\mathcal{L}'(1)} > 0.$$

Proof. Define

$$\bar{u} := \frac{\mathcal{L}_u''(1)}{\mathcal{L}''(1)} \quad \text{and} \quad \tilde{u}(\lambda) := u(\lambda) - \bar{u}.$$

Then $\mathcal{L}_{\tilde{u}}''(1) = 0$, so that

$$\mathcal{L}_u'''(1)\mathcal{L}'(1) - \mathcal{L}_u''(1)\mathcal{L}''(1) = \frac{\mathcal{L}_u''(1)}{\mathcal{L}''(1)}\Big[\mathcal{L}'''(1)\mathcal{L}'(1) - \mathcal{L}''(1)\mathcal{L}''(1)\Big] + \mathcal{L}_{\tilde{u}}'''(1)\mathcal{L}'(1). \tag{4}$$

We have to show that (4) is positive. The first term on the right-hand-side of (4) is nonnegative by Lemma 4. Next, note that $[\lambda - u^{-1}(\bar{u})]\tilde{u}(\lambda) > 0$ for all $\lambda \neq u^{-1}(\bar{u})$. This implies that

$$-\mathcal{L}_{\tilde{u}}'''(1) = \int \lambda^3 \tilde{u}(\lambda) \mathrm{e}^{-\lambda} dG(\lambda) = \int \lambda^2 [\lambda - u^{-1}(\bar{u})]\tilde{u}(\lambda) \mathrm{e}^{-\lambda} dG(\lambda) > 0.$$

Thus, the second term on the right-hand side of (4) is positive. □

A.2.2 Computing the distribution of K_n under the null

The finite-sample distribution of K_n conditional on the relevant sub-sample sizes $(M_{1,n}, M_{2,n})$ follows from a simple quantile transformation.

Proposition 4. *Under the null $\beta = 1$ and conditional on $(M_{1,n}, M_{2,n}) = (m_1, m_2)$,*

$$K_n \sim \sup_{u \in [0,1]} \left| \hat{U}_{1,m_1}(u)^2 - \hat{U}_{2,m_2}(u^2) \right|.$$

Here, $\hat{U}_{1,m_1}$ and $\hat{U}_{2,m_2}$ are independent uniform empirical distribution functions with m_1 and m_2 points of support, respectively, for given $m_1, m_2 \in \mathbb{N}$.

Proof. Along the lines of footnote 33 it is easy to show that $(\hat{H}_{1,n}, \hat{H}_{2,n}) \sim (\hat{H}^*_{1,m_1}, \hat{H}^*_{2,m_2})$ conditional on $(M_{1,n}, M_{2,n}) = (m_1, m_2)$, with $\hat{H}^*_{1,m_1}$ and $\hat{H}^*_{2,m_2}$ independent empirical distributions of random samples of sizes m_1 and m_2 from H_1 and H_2, respectively. Therefore, conditional on $(M_{1,n}, M_{2,n}) = (m_1, m_2)$

$$\begin{aligned} K_n \sim \sup_t \left| \hat{H}^*_{1,m_1}(t)^2 - \hat{H}^*_{2,m_2}(t) \right| &\sim \sup_t \left| \left(\hat{U}_{1,m_1} \circ H_1(t) \right)^2 - \hat{U}_{2,m_2} \circ H_2(t) \right| \\ &= \sup_t \left| \left(\hat{U}_{1,m_1} \circ H_1(t) \right)^2 - \hat{U}_{2,m_2} \circ H_1(t)^2 \right| \\ &= \sup_{u \in [0,1]} \left| \hat{U}_{1,m_1}(u)^2 - \hat{U}_{2,m_2}(u^2) \right|. \end{aligned}$$

□

Next, recall from Appendix A.1.4 that durations are rounded to integer days in our sample. Formally, we can only compute KS-statistics for the discretized distributions on $\{T/365, 2T/365, \ldots, T\}$. Again, the resulting statistic is not distribution-free. In the present case, in which the null hypothesis does not specify the distribution H_1 (or H_2), this complicates the computation of exact p-values. However, as argued before, the effect of the discretization is likely to be small. To check this, we have computed (sharp) bounds on the (continuous-time) KS-statistic and its p-value imposed by observations of $\hat{H}_1$ and $\hat{H}_2$ on $\{T/365, 2T/365, \ldots, T\}$. The discrete (and reported) KS-statistic provides a sharp lower bound. In the main text, we report a statistic of 0.067 with a p-value of 0.423. Note that this p-value provides an upper bound on the p-value under continuous observation. An upper bound on the KS-statistic is easily found by including distances of $\hat{H}_1(t)^2$ and $\hat{H}_2(t')$ for t and t' one day apart in the comparison. The upper bound on the statistic is 0.070. The corresponding p-value, 0.358, provides a lower bound on the p-value under continuous observation. As expected, the bounds are narrow and justify the conclusion the results are not affected by the discretization.

A.3 Results for Subsection 3.4.3 ($\widehat{\ln\beta}_n$, $\pi(\Psi)$ and $\pi(H_1)$)

A.3.1 Properties of $\widehat{\ln\beta}_n$ and $\hat{\pi}_n$

Let $\widehat{\ln\beta}_n(\Psi)$ be defined as $\widehat{\ln\beta}_n$ for the transformed durations $T^*_{1,i}$ and $T^*_{2,i} - T^*_{1,i}$. Note that $(T^*_{1,i}, T^*_{2,i} - T^*_{1,i})$ is uniformly distributed on $\{(t_1, t_2) : 0 \leq t_1 < 1, t_1 < t_2 \leq 1\}$ conditional on $N_i(T) = 2$ under the null. Using this, the asymptotic distribution of $\widehat{\ln\beta}_n(\Psi)$ under the null (and therefore $\widehat{\ln\beta}_n$ under the assumption of stationarity) is easy to derive. Let $\mathcal{N}(\mu, \sigma^2)$ denote a normal random variable with mean μ and variance σ^2. Then, we have

Proposition 5. *Under the null $\beta = 1$,*

$$\sqrt{n}\,\widehat{\ln\beta}_n(\Psi) \Longrightarrow \mathcal{N}\left(0, \frac{\pi^2}{3p_2}\right)$$

as $n \to \infty$.

Proof. Let $\beta = 1$. Then $\mathbb{E}[(\ln(T^*_{1,i}) - \ln(T^*_{2,i} - T^*_{1,i}))I(N_i(T) = 2)] = 0$ and

$$\mathbb{E}[(\ln(T^*_{1,i}) - \ln(T^*_{2,i} - T^*_{1,i}))^2 I(N_i(T) = 2)] = 2p_2 \int_0^1 \int_{t_1}^1 (\ln(t_1) - \ln(t_2 - t_1))^2 dt_2 dt_1 = p_2 \frac{\pi^2}{3}.$$

Also, $n^{-1}M_{2,n} \xrightarrow{\text{a.s.}} p_2$ as $n \to \infty$ by the law of large numbers. The result follows by the central limit theorem and Slutsky's lemma. □

The distributional properties of $\hat{\pi}_n(\Psi)$ under the null (and therefore of $\hat{\pi}_n$ under the assumption stationarity) are standard.

Proposition 6. *Under the null $\beta = 1$,*

$$\Pr\left[\hat{\pi}_n(\Psi) = \frac{i}{m}\middle| M_{2,n} = m\right] = \binom{m}{i} 2^{-m}$$

for $i = 0, \ldots, m$, and

$$\sqrt{n}\left(\hat{\pi}_n(\Psi) - \frac{1}{2}\right) \Longrightarrow \mathcal{N}\left(0, \frac{1}{4p_2}\right)$$

as $n \to \infty$.

Proof. This follows from the well-known application of the central limit theorem to the binomial distribution, $n^{-1}M_{2,n} \xrightarrow{\text{a.s.}} p_2$ as $n \to \infty$ and Slutsky's lemma. □

Finally, we can sign $\pi(\Psi) - 1/2$ under the alternative using that

$$\Pr(T^*_1 \geq T^*_2 - T^*_1 | \lambda, N(T) = 2) = \left(\frac{\beta(\lambda)}{2\beta(\lambda) + 1}\right) \frac{[2\beta(\lambda) + 1]e^{-\lambda} - 2[\beta(\lambda) + 1]e^{-\frac{1}{2}[\beta(\lambda)+1]\lambda} + e^{-\beta(\lambda)^2\lambda}}{\beta(\lambda)e^{-\lambda} - [\beta(\lambda) + 1]e^{-\beta(\lambda)\lambda} + e^{-\beta(\lambda)^2\lambda}}$$

for $\beta(\lambda) \neq 1$.

A.3.2 Adapting $\hat{\pi}_n$ to selection on $N_i(T) \geq 2$

Consider the alternative case in which we select on $N_i(T) \geq 2$. It is easy to check that the distribution of $(T_1^*, T_2^* - T_1^*)$ conditional on $N(T) \geq 2$, unlike that conditional on $N(T) = 2$, depends on the distribution G of λ under the null. However,

$$\Pr(T_1^* \geq T_2^* - T_1^* | \lambda, N_i(T) \geq 2) = \begin{cases} \left(\frac{\beta(\lambda)}{\beta(\lambda)+1}\right) \frac{[\beta(\lambda)+1]e^{-\lambda} - 2e^{-\frac{1}{2}[\beta(\lambda)+1]\lambda} + 1 - \beta(\lambda)}{\beta(\lambda)e^{-\lambda} - e^{-\beta(\lambda)\lambda} + 1 - \beta(\lambda)} & \text{if } \beta \neq 1, \text{ and} \\ \frac{1}{2} & \text{if } \beta = 1. \end{cases}$$

Under the null $\beta = 1$, we again have that $\Pr(T_1^* \geq T_2^* - T_1^* | \lambda, N(T) \geq 2) = 1/2$ is known and independent of λ. The distributional properties of the equivalent of $\hat{\pi}_n(\Psi)$ that conditions on $N_i(T) \geq 2$ are analogous to Proposition 6, with $\sum_{k=2}^{\infty} M_{k,n}$ replacing $M_{2,n}$ and $\sum_{k=2}^{\infty} p_k$ replacing p_2. Finally,

$$\Pr(T_1^* \geq T_2^* - T_1^* | \lambda, N(T) \geq 2) \begin{cases} < \frac{1}{2} & \text{if } \beta(\lambda) < 1, \text{ and} \\ > \frac{1}{2} & \text{if } \beta(\lambda) > 1 \end{cases}$$

for all $\lambda > 0$.

One attractive difference with selection on $N(T) = 2$ is that the baseline case without selection arises if we take the limit $T \to \infty$:

$$\lim_{T \to \infty} \Pr(T_1^* \geq T_2^* - T_1^* | \lambda, N(T) \geq 2) = \frac{\beta(\lambda)}{\beta(\lambda) + 1}.$$

A.3.3 Asymptotic properties of $\hat{\pi}_n(\hat{H}_{1,n})$ under the null $\beta = 1$

The main result is

Proposition 7. *Under the null $\beta = 1$,*

$$\sqrt{n}\left(\hat{\pi}_n(\hat{H}_{1,n}) - \frac{1}{2}\right) \Longrightarrow \mathcal{N}\left(0, \frac{1}{4p_2} + \frac{1}{6p_1}\right)$$

as $n \to \infty$.

Proof. Let

$$\pi : h \in \mathcal{D} \mapsto \Pr\left(2h(T_{1,i}) \geq h(T_{2,i}) | N_i(T) = 2\right), \tag{5}$$

with $\mathcal{D}$ the set of all distribution functions on $[0, T]$ concentrated on $(0, T]$. Note that π does not depend on i and that $\pi(\Psi) = 1/2$ if $\beta = 1$ (as is maintained throughout). It is convenient to write

$$\sqrt{n}\left(\hat{\pi}_n(\hat{H}_{1,n}) - \frac{1}{2}\right) = \sqrt{n}\left(\hat{\pi}_n(\hat{H}_{1,n}) - \pi(\hat{H}_{1,n})\right) + \sqrt{n}\left(\pi(\hat{H}_{1,n}) - \frac{1}{2}\right)$$

and analyze the two terms in the right-hand side. Lemmas 6 and 7 below imply that

$$\sqrt{n}\left(\hat{\pi}_n(\hat{H}_{1,n}) - \pi(\hat{H}_{1,n})\right) \Longrightarrow \mathcal{N}_1\left(0, \frac{1}{4p_2}\right)$$

and

$$\sqrt{n}\left(\pi(\hat{H}_{1,n}) - \frac{1}{2}\right) \Longrightarrow \mathcal{N}_2\left(0, \frac{1}{6p_1}\right),$$

as $n \to \infty$, with $\mathcal{N}_1\left(0, \frac{1}{4p_2}\right)$ and $\mathcal{N}_2\left(0, \frac{1}{6p_1}\right)$ independent normal random variables. The claimed result follows. $\square$

Remains to state and prove Lemmas 6 and 7.

Lemma 6. *Under the null $\beta = 1$,*

$$\sqrt{n}\left(\hat{\pi}_n(\hat{H}_{1,n}) - \pi(\hat{H}_{1,n})\right) \Longrightarrow \mathcal{N}\left(0, \frac{1}{4p_2}\right)$$

conditional on $\{\hat{H}_{1,i}\}$ almost surely as $n \to \infty$.

Proof. Denote $Y_{n,i} := I(2\hat{H}_{1,n}(T_{1,i}) \geq \hat{H}_{1,n}(T_{2,i}), N_i(T) = 2)$. Note that

$$\mathbb{E}[Y_{n,i}|\{N_j(T), \hat{H}_{1,j}\}] = I(N_i(T) = 2)\pi(\hat{H}_{1,n}) \quad \text{and}$$
$$\mathrm{var}(Y_{n,i}|\{N_j(T), \hat{H}_{1,j}\}) = I(N_i(T) = 2)\pi(\hat{H}_{1,n})\left(1 - \pi(\hat{H}_{1,n})\right)$$

almost surely, that π is continuous at Ψ, that $\sup|\hat{H}_{1,n} - \Psi| \xrightarrow{\text{a.s.}} 0$ as $n \to \infty$ (Lemma 3) and that $n^{-1}M_{2,n} \xrightarrow{\text{a.s.}} p_2$ as $n \to \infty$. It follows that for all $\epsilon > 0$

$$0 \leq \sum_{i=1}^{n} \mathbb{E}\left[\left(\sqrt{n}\frac{Y_{n,i}}{M_{2,n}}\right)^2 I\left(\sqrt{n}\frac{Y_{n,i}}{M_{2,n}} > \epsilon\right) \middle| \{N_j(T), \hat{H}_{1,j}\}\right]$$
$$\leq \left(\frac{n}{M_{2,n}}\right)^2 n^{-1}\sum_{i=1}^{n} \Pr\left(Y_{n,i} > \sqrt{n}\,\epsilon\frac{M_{2,n}}{n} \middle| \{N_j(T), \hat{H}_{1,j}\}\right) \longrightarrow 0$$

and

$$\sum_{i=1}^{n} \mathrm{var}\left(\sqrt{n}\frac{Y_{n,i}}{M_{2,n}} \middle| \{N_j(T), \hat{H}_{1,j}\}\right) = \frac{n}{M_{2,n}}\pi(\hat{H}_{1,n})\left(1 - \pi(\hat{H}_{1,n})\right)$$
$$\longrightarrow \frac{\pi(\Psi)\left(1 - \pi(\Psi)\right)}{p_2} = \frac{1}{4p_2}$$

almost surely as $n \to \infty$. As a consequence, by the Lindeberg-Feller central limit theorem (Van der Vaart, 1998, Proposition 2.27)

$$\sqrt{n}\sum_{i=1}^{n}\left(\frac{Y_{n,i}}{M_{2,n}} - \pi(\hat{H}_{1,n})\right) \Longrightarrow \mathcal{N}\left(0, \frac{1}{4p_2}\right)$$

conditional on $\{N_j(T), \hat{H}_{1,j}\}$ almost surely as $n \to \infty$. Note that the limit does not depend on $\{N_j(T), \hat{H}_{1,j}\}$ and that

$$\hat{\pi}_n(\hat{H}_{1,n}) = \sum_{i=1}^{n} \frac{Y_{n,i}}{M_{2,n}}.$$

The claimed result follows. □

Lemma 7. *Under the null $\beta = 1$,*

$$\sqrt{n}\left(\pi(\hat{H}_{1,n}) - \frac{1}{2}\right) \Longrightarrow \mathcal{N}\left(0, \frac{1}{6p_1}\right)$$

as $n \to \infty$.

Proof. We need some notation and conventions. Let $\mathcal{B}(Z)$ be the set of bounded $\mathbb{R}$-valued functions on $Z \subset \mathbb{R}$. We abbreviate $\mathcal{B}[0,T] := \mathcal{B}([0,T])$, etcetera. We equip $\mathcal{B}(Z)$ with the uniform norm and the product space $\mathcal{B}(Z)\times\mathcal{B}(Z)$ with the norm $(g_1, g_2) \in (\mathcal{B}(Z)\times\mathcal{B}(Z)) \mapsto \max\{\sup|g_1|, \sup|g_2|\}$. All subsets of $\mathcal{B}(Z)$ and $\mathcal{B}(Z)\times\mathcal{B}(Z)$ inherit the corresponding metrics. Multiplication of elements of $\mathcal{B}(Z)$ is defined as point-wise multiplication. $\mathcal{C}(Z)$ is the set of uniformly continuous functions in $\mathcal{B}(Z)$. Throughout, we take $\beta = 1$.

Below, we show that the Hadamard derivative of π at Ψ tangentially to $\mathcal{C}[0,T]$ is

$$\pi'_\Psi : u \in \mathcal{C}[0,T] \mapsto 2\int_0^{1/2} \left[2u\left(\Psi^{-1}(z)\right) - u\left(\Psi^{-1}(2z)\right)\right] dz.$$

With the facts that $\sqrt{n}\left(\hat{H}_{1,n} - \Psi\right) \Longrightarrow (1/\sqrt{p_1})\mathbb{G}_\Psi$ (Lemma 3) and that $\mathbb{G}_\Psi$ assumes values in $\mathcal{C}[0,T]$, the functional Delta method (Van der Vaart, Theorem 20.8) implies that

$$\begin{aligned}\sqrt{n}\left(\pi(\hat{H}_{1,m}) - \pi(\Psi)\right) \Longrightarrow \frac{1}{\sqrt{p_1}}\pi'_\Psi(\mathbb{G}_\Psi) &= \frac{2}{\sqrt{p_1}}\int_0^{1/2}\left[2\mathbb{G}_\Psi\left(\Psi^{-1}(z)\right) - \mathbb{G}_\Psi\left(\Psi^{-1}(2z)\right)\right] dz \\ &= \frac{2}{\sqrt{p_1}}\int_0^{1/2}\left[2\mathbb{G}_U(z) - \mathbb{G}_U(2z)\right] dz \\ &= \mathcal{N}\left(0, \frac{1}{6p_1}\right)\end{aligned}$$

as $n \to \infty$.

Remains to show that π is Hadamard differentiable at Ψ tangentially to $\mathcal{C}[0,T]$ with the derivative π'_Ψ given above. Note that the Lebesgue density of $(T_1, T_2)|(N(T) = 2)$ at (t_1, t_2) equals $2\psi(t_1)\psi(t_2)$ if $0 \le t_1 < t_2 \le T$ and 0 otherwise. Using this, we can rewrite (5) as

$$\pi(h) = 2\int_0^T\left[\int_{h^{-1}\left(\frac{1}{2}h(t)\right)}^t \psi(\tau)d\tau\right]\psi(t)dt = 1 - 2\int\left(\Psi\circ h^{-1}\circ\left(\frac{h}{2}\right)\right) d\Psi.$$

where $h^{-1}(p) = \inf\{t : h(t) \ge p\}$ is the generalized inverse of h. The functional $\pi : \mathcal{D} \to \mathbb{R}$ can be decomposed as

$$\begin{aligned}h \in \mathcal{D} &\overset{\rho}{\longmapsto} \left(\frac{h}{2}, h^{-1}\right) \in \mathcal{D}_1 \times \mathcal{I} \overset{\sigma}{\longmapsto} h^{-1}\circ\left(\frac{h}{2}\right) \in \mathcal{D}_T \\ &\overset{\tau}{\longmapsto} \Psi\circ\left(h^{-1}\circ\left(\frac{h}{2}\right)\right) \in \mathcal{B}[0,T] \overset{\upsilon}{\longmapsto} 1 - 2\int\left(\Psi\circ h^{-1}\circ\left(\frac{h}{2}\right)\right) d\Psi \in \mathbb{R}.\end{aligned}$$

Recall that $\mathcal{D} \subset \mathcal{B}[0,T]$ is the set of distribution functions on $[0,T]$ concentrated on $(0,T]$. Also, $\mathcal{D}_1 \subset \mathcal{B}[0,T]$ and $\mathcal{D}_T \subset \mathcal{B}[0,T]$ are the sets of, respectively, $[0,1)$-valued and $[0,T]$-valued functions on $[0,T]$ and $\mathcal{I} \subset \mathcal{B}[0,1)$ is the set of $[0,T]$-valued functions on $[0,1)$.

We first show that each of the maps ρ, σ, τ and υ is (tangentially) Hadamard differentiable and then derive π'_Ψ by the chain rule.

(i). *Hadamard derivative of $\rho : \mathcal{D} \to \mathcal{D}_1 \times \mathcal{I}$ at Ψ tangentially to $\mathcal{C}[0,T]$*

By Van der Vaart (1998), Lemma 21.4, the inverse map $h \in \mathcal{D} \subset \mathcal{B}[0,T] \mapsto h^{-1} \in \mathcal{I}$ is Hadamard differentiable at Ψ tangentially to $\mathcal{C}[0,T]$ with derivative $u \in \mathcal{C}[0,T] \mapsto -(u/\psi)\circ\Psi^{-1}$. Thus, the Hadamard derivative of ρ at Ψ tangentially to $\mathcal{C}[0,T]$ is

$$\rho'_\Psi : u \in \mathcal{C}[0,T] \mapsto \left(\frac{u}{2}, -\left(\frac{u}{\psi}\right)\circ\Psi^{-1}\right) \in \mathcal{B}[0,T] \times \mathcal{C}[0,1).$$

(ii). *Hadamard derivative of* $\sigma : \mathcal{D}_1 \times \mathcal{I} \to \mathcal{D}_T$ *at* $\rho(\Psi) = \left(\Psi/2, \Psi^{-1}\right)$ *tangentially to* $\mathcal{B}[0,T] \times \mathcal{C}[0,1)$

By the assumption that Ψ is continuously differentiable with positive derivative ϕ on $[0,T]$, Ψ^{-1} is uniformly differentiable on $[0,1)$ with uniformly bounded derivative $(1/\psi) \circ \Psi^{-1}$. Thus, by Van der Vaart and Wellner (1996), Lemma 3.9.27, the Hadamard derivative of σ at $\left(\Psi/2, \Psi^{-1}\right)$ tangentially to $\mathcal{B}[0,T] \times \mathcal{C}[0,1)$ is[34]

$$\sigma'_{\rho(\Psi)} : (u,v) \in \mathcal{B}[0,T] \times \mathcal{C}[0,1) \mapsto v \circ \left(\frac{\Psi}{2}\right) + \frac{u}{\psi \circ \Psi^{-1} \circ \left(\frac{\Psi}{2}\right)} \in \mathcal{B}[0,T].$$

(iii). *Hadamard derivative of* $\tau : \mathcal{D}_T \to \mathcal{B}[0,T]$ *at* $\sigma(\rho(\Psi)) = \Psi^{-1} \circ (\Psi/2)$

Let $u_q \to u \in \mathcal{B}[0,T]$ uniformly as $q \downarrow 0$ and $\sigma(\rho(\Psi)) + qu_q \in \mathcal{D}_T$ for all q. Abbreviate $s := \sigma(\rho(\Psi))$ and consider

$$\begin{aligned}
0 &\le \left| \frac{\tau(s+qu_q) - \tau(s)}{q} - (\psi \circ s)u \right| \\
&= \left| \frac{\Psi \circ (s+qu_q) - \Psi \circ s}{q} - (\psi \circ s)u_q + (\psi \circ s)\left(u_q - u\right) \right| \\
&\le \left| \frac{\Psi \circ (s+qu_q) - \Psi \circ s}{q} - (\psi \circ s)u_q \right| + (\psi \circ s)\left|u_q - u\right|
\end{aligned}$$

The first term in the last line converges uniformly to 0 as $q \downarrow 0$ by the uniform differentiability of Ψ. The second term converges uniformly to 0 because ψ is uniformly bounded. Thus, the right-hand side of the first line converges uniformly to 0 as $q \downarrow 0$ and the Hadamard derivative of $\tau : \mathcal{D}_T \to \mathcal{B}[0,T]$ at $\sigma(\rho(\Psi)) = \Psi^{-1} \circ (\Psi/2)$ is

$$\tau'_{\sigma(\rho(\Psi))} : u \in \mathcal{B}[0,T] \mapsto \left(\psi \circ \Psi^{-1} \circ \left(\frac{\Psi}{2}\right)\right) u \in \mathcal{B}[0,T].$$

(iv). *Hadamard derivative of* $\upsilon : \mathcal{B}[0,T] \to \mathbb{R}$ *at* $\tau(\sigma(\rho(\Psi))) = \Psi/2$

Let $u_q \to u \in \mathcal{B}[0,T]$ uniformly as $q \downarrow 0$ and $\tau(\sigma(\rho(\Psi))) + qu_q \in \mathcal{B}[0,T]$ for all q. Then,

$$\frac{\upsilon(\Psi/2 + qu_q) - \upsilon(\Psi/2)}{q} = -2\int u_q d\Psi \longrightarrow -2\int u d\Psi$$

as $q \downarrow 0$ because the sequence u_q is uniformly bounded. So, the Hadamard derivative of υ : $\mathcal{B}[0,T] \to \mathbb{R}$ at $\tau(\sigma(\rho(\Psi))) = \Psi/2$ is

$$\upsilon'_{\tau(\sigma(\rho(\Psi)))} : u \in \mathcal{B}[0,T] \mapsto -2\int u d\Psi.$$

[34]It may be helpful here to construct the proof of this lemma for our special case. Let $(u_q, v_q) \to (u,v) \in \mathcal{B}[0,T] \times \mathcal{C}[0,1)$ uniformly as $q \downarrow 0$ and $\rho(\Psi) + q(u_q, v_q) \in \mathcal{D}_1 \times \mathcal{I}$ for all q. We have that

$$\begin{aligned}
0 &\le \left| \frac{\left(\Psi^{-1} + qv_q\right) \circ \left(\frac{\Psi}{2} + qu_q\right) - \Psi^{-1} \circ \left(\frac{\Psi}{2}\right)}{q} - v \circ \left(\frac{\Psi}{2}\right) - \frac{u}{\psi \circ \Psi^{-1} \circ \left(\frac{\Psi}{2}\right)} \right| \\
&\le \left| \frac{\Psi^{-1} \circ \left(\frac{\Psi}{2} + qu_q\right) - \Psi^{-1} \circ \left(\frac{\Psi}{2}\right)}{q} - \frac{u_q}{\psi \circ \Psi^{-1} \circ \left(\frac{\Psi}{2}\right)} \right| + \left| \frac{u_q - u}{\psi \circ \Psi^{-1} \circ \left(\frac{\Psi}{2}\right)} \right| \\
&\quad + \left| v \circ \left(\frac{\Psi}{2} + qu_q\right) - v \circ \left(\frac{\Psi}{2}\right) \right| + \left| (v_q - v) \circ \left(\frac{\Psi}{2} + qu_q\right) \right|
\end{aligned}$$

The first term in the right-hand side converges (uniformly) to 0 because of the uniform differentiability of Ψ^{-1}. The second term converges to 0 because $\sup |u_q - u| \to 0$ and the denominator is bounded away from 0. The third term converges to 0 because v is uniformly continuous. The fourth term converges to 0 because $\sup |v_q - v| \to 0$.

By the chain rule (Van der Vaart, 1998, Proposition 20.9), the Hadamard derivative of $\pi = \upsilon \circ \tau \circ \sigma \circ \rho$ at Ψ tangentially to $\mathcal{C}[0,T]$ is $\upsilon'_{\tau(\sigma(\rho(\Psi)))} \circ \tau'_{\sigma(\rho(\Psi))} \circ \sigma'_{\rho(\Psi)} \circ \rho'_{\Psi}$ (note that the tangent spaces are properly lined up). Substitution of the derivative maps derived in (i)–(iv) gives the desired result. □

A.3.4 The behavior of $\pi(H_1)$ under local alternatives to $\beta = 1$

This appendix provides details on the directional derivatives of $\pi(\Psi;\beta)$ and $\pi(H_1(\beta);\beta)$ at $\beta = 1$ in the direction u.

First consider $\pi(\Psi;\beta)$. We have earlier seen that the Lebesgue density of $(T_1,T_2)|(N(T)=2)$ at (t_1,t_2) equals $2\psi(t_1)\psi(t_2)$ if $0 \le t_1 < t_2$ and 0 otherwise if $\beta = 1$. For $\beta(\lambda) \neq 1$, the density of $(T_1,T_2)|(\lambda, N(T)=2)$ at (t_1,t_2) equals

$$\frac{\lambda^2\beta(\lambda)\psi(t_1)\psi(t_2)\mathrm{e}^{-\lambda[1-\beta(\lambda)]\Psi(t_1)-\lambda\beta(\lambda)[1-\beta(\lambda)]\Psi(t_2)-\lambda\beta(\lambda)^2}}{\int_0^T\int_0^{\tau_2}\lambda^2\beta(\lambda)\psi(\tau_1)\psi(\tau_2)\mathrm{e}^{-\lambda[1-\beta(\lambda)]\Psi(\tau_1)-\lambda\beta(\lambda)[1-\beta(\lambda)]\Psi(\tau_2)-\lambda\beta(\lambda)^2}d\tau_1 d\tau_2}$$

for $0 \le t_1 < t_2 \le T$ and 0 elsewhere. Thus, for $\beta = 1 + qu$ and $q \neq 0$ we have that

$$\pi(\Psi;\beta) = \frac{\int\int_0^1\int_{\tau_2/2}^{\tau_2}\lambda^2[1+qu(\lambda)]\mathrm{e}^{\lambda qu(\lambda)\tau_1+\lambda[1+qu(\lambda)]qu(\lambda)\tau_2-\lambda[1+qu(\lambda)]^2}d\tau_1 d\tau_2 dG(\lambda)}{\int\int_0^1\int_0^{\tau_2}\lambda^2[1+qu(\lambda)]\mathrm{e}^{\lambda qu(\lambda)\tau_1+\lambda[1+qu(\lambda)]qu(\lambda)\tau_2-\lambda[1+qu(\lambda)]^2}d\tau_1 d\tau_2 dG(\lambda)}$$

Using that

$$\frac{d}{dq}\Big[\int\int_0^1\int_{\tau_2/2}^{\tau_2}\lambda^2[1+qu(\lambda)]\mathrm{e}^{\lambda qu(\lambda)\tau_1+\lambda[1+qu(\lambda)]qu(\lambda)\tau_2-\lambda[1+qu(\lambda)]^2}d\tau_1 d\tau_2 dG(\lambda)\Big]_{q=0} = \frac{5}{24}\mathcal{L}_u'''(1)+\frac{1}{4}\mathcal{L}_u''(1)$$

and that

$$\frac{d}{dq}\Big[\int\int_0^1\int_0^{\tau_2}\lambda^2[1+qu(\lambda)]\mathrm{e}^{\lambda qu(\lambda)\tau_1+\lambda[1+qu(\lambda)]qu(\lambda)\tau_2-\lambda[1+qu(\lambda)]^2}d\tau_1 d\tau_2 dG(\lambda)\Big]_{q=0} = \frac{1}{2}\mathcal{L}_u'''(1)+\frac{1}{2}\mathcal{L}_u''(1)$$

we find that

$$\frac{d}{dq}\Big[\pi(\Psi;1+qu)\Big]_{q=0} = \frac{1}{12}\frac{-\mathcal{L}_u'''(1)}{\mathcal{L}''(1)}. \tag{G2}$$

Next, consider $\pi(H_1(\beta);\beta)$. The analysis in A.2.1 implies that

$$\frac{d}{dq}\Big[H_1(1+qu)\Big]_{q=0} = -\frac{1}{2}\frac{\mathcal{L}_u''(1)}{-\mathcal{L}'(1)}\Psi(1-\Psi)$$

so that

$$\frac{d}{dq}\Big[\pi(H_1(1+qu);1)\Big]_{q=0} = \pi'_\Psi\left(-\frac{1}{2}\frac{\mathcal{L}_u''(1)}{-\mathcal{L}'(1)}\Psi(1-\Psi)\right) = -\frac{1}{12}\frac{\mathcal{L}_u''(1)}{-\mathcal{L}'(1)}.$$

Thus, it follows that

$$\begin{aligned}\frac{d}{dq}\Big[\pi(H_1(1+qu);1+qu)\Big]_{q=0} &= \frac{d}{dq}\Big[\pi(\Psi;1+qu)\Big]_{q=0} + \frac{d}{dq}\Big[\pi(H_1(1+qu);1)\Big]_{q=0}\\ &= \frac{1}{12}\left[\frac{-\mathcal{L}_u'''(1)}{\mathcal{L}''(1)} - \frac{\mathcal{L}_u''(1)}{-\mathcal{L}'(1)}\right].\end{aligned} \tag{G3}$$

B Identification and estimation

This appendix provides results for Subsection 3.5. Note that the identification and estimation analyses in this subsection assume that β is homogeneous, as in ($M^\dagger$).

B.1 Results for Subsection 3.5.1 (Identification)

Proof of Proposition 2. Denote the sub-density of $(T_1, \ldots, T_k)$ on $N(T) = k$ by f_k (i.e., $f_1(t) = d\Pr(T_1 \leq t, N(T) = 1)/dt$, etcetera). We first show that the cases $\beta = 1$ and $\beta \neq 1$ can be distinguished from data on f_1 and f_2. With the normalization $\Psi(T) = 1$, we have that

$$\frac{f_2(t_1, t_2)}{f_1(t_1)f_1(t_2)} = \beta \frac{\mathcal{L}''\left((1-\beta)\left(\Psi(t_1) + \beta\Psi(t_2)\right) + \beta^2\right)}{\mathcal{L}'\left((1-\beta)\Psi(t_1) + \beta\right)\mathcal{L}'\left((1-\beta)\Psi(t_2) + \beta\right)}.$$

Straightforward calculations show that

$$\left[\frac{\partial \ln\left(\frac{f_2(t_1,t_2)}{f_1(t_1)f_1(t_2)}\right)}{\partial t_1} - \frac{\partial \ln\left(\frac{f_2(t_1,t_2)}{f_1(t_1)f_1(t_2)}\right)}{\partial t_2}\right]_{t_1=t_2=t} = (1-\beta)^2\psi(t)\frac{\mathcal{L}'''\left((1-\beta^2)\Psi(t) + \beta^2\right)}{\mathcal{L}''\left((1-\beta^2)\Psi(t) + \beta^2\right)},$$

which equals 0 (for all t) if and only if $\beta = 1$. This establishes whether $\beta = 1$.

In the case that $\beta \neq 1$, it remains to distinguish between $\beta < 1$ and $\beta > 1$. To this end, note that the claim intensity at t conditional on $(T_1 = t_1, T_2 \geq t)$, $t > t_1$, is given by

$$\theta(t|T_1 = t_1, T_2 \geq t) = \beta\psi(t)\frac{\mathcal{L}''\left((1-\beta)\Psi(t_1) + \beta\Psi(t)\right)}{-\mathcal{L}'\left((1-\beta)\Psi(t_1) + \beta\Psi(t)\right)}.$$

Lemma 4 in Appendix A.2 implies that the function $\mathcal{L}''/(-\mathcal{L}')$ is trivial if λ either is degenerate or has two points of support of which one is 0. Otherwise, $\mathcal{L}''/(-\mathcal{L}')$ is (strictly) decreasing. Clearly, if $\theta(t|T_1 = t_1, T_2 \geq t)$ is decreasing in t_1, we know that $\beta < 1$. Similarly, if $\theta(t|T_1 = t_1, T_2 \geq t)$ is increasing in t_1 then we can conclude that $\beta > 1$. If $\theta(t|T_1 = t_1, T_2 \geq t)$ is constant in t_1, we know (because $\beta \neq 1$) that λ either is degenerate or has two points of support of which one is 0. In that case,

$$\mathcal{L}(s) = \Pr(\lambda = 0) + (1 - \Pr(\lambda = 0))\exp(-\mathbb{E}[\lambda|\lambda > 0]s)$$

and $\theta(t|T_1 = t_1, T_2 \geq t) = \mathbb{E}[\lambda|\lambda > 0]\beta\psi(t)$. With the normalization $\Psi(T) = 1$, this identifies $\mathbb{E}[\lambda|\lambda > 0]\beta$ and Ψ. Then,

$$\Pr(T_1 > t) = \mathcal{L}\left(\Psi(t)\right) = \Pr(\lambda = 0) + (1 - \Pr(\lambda = 0))\exp(-\mathbb{E}[\lambda|\lambda > 0]\Psi(t))$$

identifies $\Pr(\lambda = 0)$ and $\mathbb{E}[\lambda|\lambda > 0]$ and therewith β. □

Note that the proof only uses data on first and second claim times (that is, the sub-distributions of T_1 on $N(T) = 1$ and T_2 on $N(T) = 2$ and the claim intensities at 0 and 1 claim). A central role is played by the way $\theta(t|T_1 = t_1, T_2 \geq t)$ depends on t_1. Conditional on λ however, the claim intensity only depends on the occurrence, and not the timing, of past claims. Thus, the dependence of $\theta(t|T_1 = t_1, T_2 \geq t)$ on t_1 works by way of the heterogeneity. This explains that a special role is played by the cases that λ is degenerate and that λ has two points of support of which one is 0. In either case, λ is degenerate at the non-zero point of support $\mathbb{E}[\lambda|\lambda > 0]$ conditional on past occurrence of a claim and there is no heterogeneity conditional on $(T_1 = t_1, T_2 \geq t)$. We have seen these cases appearing in the analyses of the behavior of the two general tests in Subsections 3.4.2 and 3.4.3 under homogeneous-β local alternatives. Note that these alternatives fit the special model ($M^\dagger$) that we are studying here.

Next, consider the key identifying equation (I). Recall that $q_0(t) = \mathcal{L}(\Psi(t))$ and note that

$$q_1(t) = \frac{\mathcal{L}(\beta\Psi(t)) - \mathcal{L}(\Psi(t))}{1-\beta} \quad \text{and} \quad q_2(t) = \frac{\beta\mathcal{L}(\Psi(t)) - (1+\beta)\mathcal{L}(\beta\Psi(t)) + \mathcal{L}(\beta^2\Psi(t))}{(1-\beta^2)(1-\beta)}.$$

if $\beta \neq 1$. Also, define $\tilde{q}(t) := (1-\beta)q_1(t) + q_0(t) = \mathcal{L}(\beta\Psi(t))$. If $\beta \neq 1$ then

$$q_1\{q_0^{-1}[\tilde{q}(t)]\} = \frac{\mathcal{L}(\beta^2\Psi(t)) - \mathcal{L}(\beta\Psi(t))}{1-\beta},$$

which indeed equals $\beta q_1(t) + (1-\beta^2)q_2(t)$.

Proof of Proposition 3. The case $\beta = 1$ has been covered in Subsection 3.4.1. So, consider the case that $\beta \neq 1$. If we know β we can compute the function $\tilde{q}$ defined above. The remainder of the proof closely follows Kortram et al. (1995), in particular a version thereof by Abbring (2002) that can directly be applied to our finite-support setup here. Suppose that $\beta > 1$ (the case $\beta < 1$ is similar). We have that, for any $t \in [0,T]$, $y := \tilde{q}(t) = \mathcal{L}(\beta\Psi(t)) \Leftrightarrow \Psi(t) = \beta^{-1}\mathcal{L}^{-1}(y)$. Also, $t = \tilde{q}^{-1}(y)$, so that

$$q_0(\tilde{q}^{-1}(y)) = q_0(t) = \mathcal{L}(\Psi(t)) = \mathcal{L}(\beta^{-1}\mathcal{L}^{-1}(y))$$

By induction, it follows that, for $y \in [\tilde{q}(T), 1]$,

$$(q_0 \circ \tilde{q}^{-1})^{(n)}(y) = \mathcal{L}(\beta^{-n}\mathcal{L}^{-1}(y)), \tag{6}$$

where $q_0 \circ \tilde{q}^{-1}$ is the composition of q_0 with $\tilde{q}$, and superscript (n) denotes the n-fold composition. Now, l'Hôspital's rule gives

$$\mathbb{E}[\lambda]\mathcal{L}^{-1}(y) = \lim_{n\to\infty} \frac{1 - \mathcal{L}(\beta^{-n}\mathcal{L}^{-1}(y))}{\beta^{-n}} = \lim_{n\to\infty} \frac{\left(1 - (q_0 \circ \tilde{q}^{-1})^{(n)}(y)\right)}{\beta^{-n}} \tag{7}$$

on $[\tilde{q}(T), 1]$. Evaluating (7) at $q_0(T) \in [\tilde{q}(T), 1]$ gives $\mathbb{E}[\lambda]$, because $\mathcal{L}^{-1}(q_0(T)) = \Psi(T) = 1$. So, $\mathcal{L}^{-1}$ is uniquely determined on $[\tilde{q}(T), 1]$. We then have that $\Psi(t) = \mathcal{L}^{-1}(q_0(t)) = \beta^{-1}\mathcal{L}^{-1}(\tilde{q}(t))$ on $[0,T]$. From $\mathcal{L}^{-1}$ on $[\tilde{q}(T), 1]$ we can identify $\mathcal{L}$ on $[0, \mathcal{L}^{-1}(\tilde{q}(T)] = [0, \beta\Psi(T)]$. Finally, $\mathcal{L}$ can be analytically extended to $[0, \infty)$. □

B.2 Results for Subsection 3.5.2 (Estimation)

In this appendix we construct the likelihood on which the estimates in Subsection 3.5.2 are based. Choose some marginal density $g(\cdot;\xi)$ of λ_i and some contract-time function $\psi(\cdot;\alpha)$. As before, we define $\Psi(t;\alpha) := \int_0^t \psi(u;\alpha)du$ and make sure that the normalization $\Psi(T;\alpha) = 1$ is satisfied. Both ξ and α are finite-dimensional parameter vectors.

Consider the likelihood contribution L_i of contract i. For contract i, we observe a claim history $N_i[0,T]$ (we ignore the discretization of time in days here). First consider the likelihood contribution for the case that we observe λ_i. Now, recall that $T_{k,i}$ is the time of the k-th claim and let $T_{0,i} := 0$. For contract i, we observe the number of claims $N_i(T)$ in the contract year, and, if $N_i(T) \geq 1$, the times $T_{1,i}, \ldots, T_{N_i(T),i}$ of these claims. Straightforward calculations show that this "full information" likelihood contribution of contract i is

$$L_i(\beta, \alpha, \xi; N_i[0,T], \lambda_i) = \left[\prod_{k=1}^{N_i(T)} \lambda_i\beta^{k-1}\psi(T_{k,i};\alpha)\mathrm{e}^{-\lambda_i\beta^{k-1}[\Psi(T_{k,i};\alpha) - \Psi(T_{k-1,i};\alpha)]}\right]$$
$$\times\, \mathrm{e}^{-\lambda_i\beta^{N_i(T)}\left[1-\Psi(T_{N_i(T),i};\alpha)\right]}g(\lambda_i;\xi)$$

if $N_i(T) \geq 1$, and

$$L_i(\beta, \alpha, \xi; N_i[0,T], X_i, \lambda_i) = \mathrm{e}^{-\lambda_i} g(\lambda_i; \xi)$$

if $N_i(T) = 0$. Thus, the marginal likelihood contribution for the case we do not observe λ_i is

$$\begin{aligned} L_i(\beta, \alpha, \xi; N_i[0,T]) = & \left[\prod_{k=1}^{N_i(T)} \beta^{k-1} \psi(T_{k,i}; \alpha) \right] \\ & \times (-1)^{N_i(T)} \mathcal{L}^{(N_i(T))} \Big[\beta^{N_i(T)} \left[1 - \Psi(T_{N_i(T),i}; \alpha) \right] \\ & \qquad + \sum_{k=1}^{N_i(T)} \beta^{k-1} \left[\Psi(T_{k,i}; \alpha) - \Psi(T_{k-1,i}; \alpha) \right]; \xi \Big] \end{aligned}$$

if $N_i(T) \geq 1$, and

$$L_i(\beta, \alpha, \xi; N_i[0,T]) = \mathcal{L}(1; \xi)$$

if $N_i(T) = 0$. Here, $\mathcal{L}^{(k)}(\cdot; \xi)$ is the k-th derivative of the Laplace transform of $g(\cdot; \xi)$.

References

Abbring, J.H. (2002), “Stayers versus defecting movers: A note on the identification of defective duration models”, *Economics Letters*, 74, 327–331.

Araujo, A. and H. Moreira (2002), “Adverse selection without the single crossing condition”, Mimeo, Getulio Vargas Foundation.

Bates, G.E. and J. Neyman (1952), “Contributions to the theory of accident proneness II: True or false contagion”, *University of California Publications in Statistics*, 1, 255–275.

Cardon, J., and I. Hendel (1998), “Asymmetric information in health insurance: Evidence from the National Health Expenditure Survey”, Mimeo, Princeton University.

Chamberlain, G. (1985), “Heterogeneity, omitted variable bias, and duration dependence”, in J.J. Heckman and B. Singer, editors, *Longitudinal Analysis of Labor Market Data*, Cambridge University Press, Cambridge.

Chassagnon, A. and P.A. Chiappori, P.A. (2000), “Insurance under moral hazard and adverse selection: the case of pure competition”, mimeo, DELTA.

Chiappori, P.A. (2000), “Econometric models of insurance under asymmetric information”, in G. Dionne, ed., *Handbook of Insurance*, North Holland.

Chiappori, P.A., F. Durand and P.Y. Geoffard (1998), “Moral hazard and the demand for physician services: First lessons from a French natural experiment”, *European Economic Review*, 42, 499–511.

Chiappori, P.A., P.Y. Geoffard and E. Kyriadizou (1998), “Cost of time, moral hazard, and the demand for physician services”, Mimeo, University of Chicago.

Chiappori, P.A., B. Jullien, B. Salanié and F. Salanié (2001), “Asymmetric information in insurance: A robust characterization”, Mimeo, University of Chicago.

Chiappori, P.A., I. Macho, P. Rey and B. Salanié (1994), ”Repeated moral hazard: Memory, commitment, and the access to credit markets”, *European Economic Review*, 1527–1553.

Chiappori, P.A. and B. Salanié (1996), “Empirical contract theory: The case of insurance data”, *European Economic Review*, 41, 943–951.

Chiappori, P.A. and B. Salanié (2000), “Testing for asymmetric information in insurance markets”, *Journal of Political Economy*, 108, 56–78.

Chiappori, P.A., B. Salanié and J. Valentin (1999), “Early starters vs late beginners”, *Journal of Political Economy*, 107, 731–760.

Dionne, G., and N. Doherty (1994), ”Adverse selection, commitment and renegotiation: Extension to and evidence from insurance markets’, *Journal of Political Economy*, 102, 210–235.

Dionne, G., C. Gouriéroux and C. Vanasse (1999), "Evidence of adverse selection in automobile insurance markets", in G. Dionne and C. Laberge-Nadeau, editors, *Automobile Insurance: Road Safety, Insurance Fraud and Regulation*, Kluwer Academic Publishers, Boston, 13–46.

Dionne, G., C. Gouriéroux and C. Vanasse (2001), "Testing for evidence of adverse selection in the automobile insurance market: A comment", *Journal of Political Economy*, 109, 444–453.

Dionne, G., M. Maurice, J. Pinquet and C. Vanasse (2001), "The role of memory in long-term contracting: Empirical evidence in automobile insurance", Mimeo, HEC, Montreal.

Dionne, G. and C. Vanasse (1992), "Automobile insurance ratemaking in the presence of asymmetrical information", *Journal of Applied Econometrics*, 7, 149–165.

Elbers, C. and G. Ridder (1982), "True and spurious duration dependence: The identifiability of the proportional hazard model", *Review of Economic Studies*, 64, 403–409.

Feller, W. (1943), "On a general class of 'contagious' distributions", *Annals of Mathematical Statistics*, 14, 389–400.

Finkelstein, A. and J. Poterba (2002), "Adverse selection in insurance markets: Policyholder evidence from the U.K. annuity market ", Mimeo, Harvard University.

Fudenberg, D., B. Holmstrom and P. Milgrom (1990), "Short-term contracts and long-term agency relationship", *Journal of Economic Theory*, 51, 1–31.

de Garidel, T. (1997), "Pareto improving asymmetric information in a dynamic insurance market", Discussion Paper 266, LSE, London.

Heckman, J.J. (1981), "Heterogeneity and state dependence", in S. Rosen, ed., *Studies in Labor Markets*, University of Chicago Press.

Heckman, J.J. and G.J. Borjas (1980), "Does unemployment cause future unemployment? Definitions, questions and answers from a continuous time model of heterogeneity and state dependence", *Economica*, 47, 247–283.

Heckman, J.J., R. Robb and J.R. Walker (1990), "Testing the mixture of exponentials hypothesis and estimating the mixing distribution by the method of moments", *Journal of the American Statistical Association*, 85, 582–589.

Hendel, I. and A. Lizzeri (1999), "The role of commitment in dynamic contracts: Evidence from life insurance", Working Paper, Princeton University.

Holly, A., L. Gardiol, G. Domenighetti and B. Bisig (1998), "An econometric model of health care utilization and health insurance in Switzerland", *European Economic Review*, 42, 513–522.

Holt, J.D. and R.L. Prentice (1974), "Survival analysis in twin studies and matched pair experiments", *Biometrika*, 61, 17–30.

Jullien, B., B. Salanié and F. Salanié (1999), "Should more risk-averse agents exert more effort?", *Geneva Papers on Risk and Insurance Theory*, 24, 19–28.

Kortram, R.A., A.J. Lenstra, G. Ridder and A.C.M. van Rooij (1995), "Constructive identification of the mixed proportional hazards model", *Statistica Neerlandica*, 49, 269–281.

Manning, W., J. Newhouse, N. Duan, E. Keeler and A. Leibowitz (1987), "Health insurance and the demand for medical care: Evidence from a randomized experiment", *American Economic Review*, 77, 251–277.

Mossin, J. (1968), "Aspects of rational insurance purchasing", *Journal of Political Economy*, 79, 553–568.

Puelz, R. and A. Snow (1994), "Evidence on adverse selection: Equilibrium signalling and cross-subsidization in the insurance market", *Journal of Political Economy*, 102, 236–257.

Richaudeau, D. (1999), "Automobile insurance contracts and risk of accident: An empirical test using French individual data", *The Geneva Papers on Risk and Insurance Theory,* 24, 97–114.

Ridder, G. (1990), "The non-parametric identification of generalized accelerated failure-time models", *Review of Economic Studies*, 57, 167–182.

Ridder, G. and I. Tunalı (1999), "Stratified partial likelihood estimation", *Journal of Econometrics*, 92, 193–232.

Rochet, J.C. and L. Stole (2000), "The economics of multidimensional screening", in *Advances in Economic Theory, 7th World Congress of the Econometric Society.*

Shohat, J.A. and J.D. Tamarkin (1943), *The Problem of Moments*, American Mathematical Society, New York.

Van den Berg, G.J. (2001), "Duration models: Specification, identification, and multiple durations", in J.J. Heckman and E. Leamer, editors, *Handbook of Econometrics, Volume V*, North Holland, Amsterdam.

Van der Vaart, A.W. (1998), *Asymptotic Statistics*, Cambridge University Press, Cambridge, UK.

Van der Vaart, A.W. and J.A. Wellner (1996), *Weak Convergence and Empirical Processes*, Springer, New York.

Widder, D.V. (1946), *The Laplace Transform*, Princeton University Press, Princeton, NJ.

Table and figures

Table 1: Maximum-likelihood estimates (discrete heterogeneity; 12 time intervals ψ)

occurrence dependence		
β	0.974	(0.677)
unobserved heterogeneity		
λ^a	0.052	(0.006)
λ^b	0.310	(0.427)
$\Pr(\lambda = \lambda^a)$	0.939	(0.104)
$\Pr(\lambda = \lambda^b)$	0.061	(0.104)
contract time (piecewise-constant ψ)		
Wald statistic $\psi \equiv 1$	15.617	
degrees of freedom	11	
p-value	0.156	
number of observations by number of claims		
$M_{0,n}$ (no claims)	74566	
$M_{1,n}$ (1 claim)	4831	
$M_{2,n}$ (2 claims)	270	
$M_{3,n}$ (3 claims)	15	
$M_{4,n}$ (4 claims)	2	
log-likelihood	-3536.16	

Figure 1: Empirical distribution function $\hat{H}_{1,n}$ of $T_1|N(1) = 1$, $\hat{H}^2_{1,n}$, and empirical distribution function $\hat{H}_{2,n}$ of $T_2|N(1) = 2$ (top), and kernel estimate of the density of $T_1|N(1) = 1$ (h_1) and maximum-likelihood estimate of ψ (bottom)

Note: Based on sample and estimates Table 1. For the bottom graph, an Epanechnikov kernel with bandwidth 0.05 is used (see Appendix A.1).

Figure 2: Maximimum-likelihood estimate Ψ

Note: Based on sample and estimates Table 1.