Unlucky or Risky? Unobserved Heterogeneity and Experience Rating in Insurance Markets*

Levon Barseghyan
Cornell University

Francesca Molinari
Cornell University

Darcy Steeg Morris
U.S. Census Bureau

Joshua C. Teitelbaum
Georgetown University

Draft: September 30, 2014
Prior version: November 14, 2012

Abstract

We investigate whether an insured's claims experience contains valuable information about its latent risk type. Using data on households' claims histories in auto and home insurance, we estimate the variance-covariance matrix of unobserved heterogeneity and utilize the estimates to update a priori predictions about the households' claim risk. The estimates reveal that unobserved heterogeneity is positively correlated across coverages. We then explore how households' demand for insurance would respond to experience rating under different theories of risky choice, and we discuss what our findings imply about the economic consequences of legal restrictions on experience rating.

(*JEL* C58, D12, D82, G22, K23)

---

*Corresponding author: Joshua C. Teitelbaum, Georgetown University Law Center, 600 New Jersey Avenue NW, Washington, DC 20001 (jct48@law.georgetown.edu). We acknowledge financial support from National Science Foundation grant SES-1031136. Molinari also acknowledges financial support from NSF grant SES-0922330. This paper is released to inform interested parties of research and to encourage discussion. The views expressed are those of the authors and not necessarily those of the U.S. Census Bureau.

1 Introduction

In many insurance markets, there are variables that affect an insured's claim risk but are not observable by the insurer.[1] In other words, there is unobserved heterogeneity in claim risk. Of course, even if there is unobserved heterogeneity in claim risk at the time the insurer underwrites and rates an insured's policy, the insurer subsequently receives signals about the insured's latent risk type. In particular, the insurer observes the insured's claims experience. A key question is whether and to what extent an insured's claims experience contains valuable information about its claim risk. Do claims signify that an insured is risky or just unlucky? If the former, then the insurer can update its prior belief about the insured's risk type. In particular, the insurer can update its prior prediction about an insured's claim risk by conditioning on the insured's claims experience. The insurer can then use its posterior prediction to adjust—or experience rate—the insured's premium, subject of course to any applicable legal restrictions on experience rating.

An important related question is whether and to what extent an insured's claims experience in one line of insurance contains valuable information about its claim risk in another line of insurance. For example, do claims in automobile insurance signify that an insured is a risky homeowner (and vice versa)? If so, then the insurer can update its prior prediction about an insured's claim risk in one line of insurance by conditioning on the insured's claims experience in another line of insurance. The insurer can then experience rate the insured's premiums across lines of insurance, subject again to any legal restrictions.

In this paper, we utilize data on claims in automobile and homeowners insurance to explore the information value of insureds' claims experience. The data comprise an unbalanced panel of $62,425$ households who held auto and home policies between 1998 and 2006. Among other things, the data record the number of claims filed by each household in three lines of coverage: auto col-

[1] Of course, these variables may not be observable by the insured either. Alternatively, there may be variables that are observable by the insurer (and perhaps also by the insured) but that the insurer is prohibited from using when it underwrites and rates the insured's policy (Salanié 1997; Avraham et al. 2014).

lision, auto comprehensive, and home all perils. In addition, the data contain detailed information about the households and their policies.

As a first step, we use the data to estimate the variance-covariance matrix $\mathbf{\Sigma}$ of unobserved heterogeneity in claim risk, as well as the households' a priori claim rates based on observables. We model households' claim counts using a Poisson mixture model with correlated random effects. To estimate the model, we take a moments-based approach that uses generalized estimating equations based on marginal moments (Morris 2012). Unlike the standard approach—maximum likelihood estimation of a parametric mixture of Poisson distributions—our estimation approach is semiparametric and unconstrained with respect to the parameters of the mixing distribution (Pinquet 2013). Among other things, the estimates reveal that unobserved heterogeneity is positively correlated across lines of coverage—0.29 between auto collision and home, 0.56 between auto comprehensive and home, and 0.66 between auto collision and auto comprehensive—suggesting that there is a domain-general component to risk type.

We then demonstrate the value of the information contained in $\widehat{\mathbf{\Sigma}}$—and, by implication, the value of the information contained in the households' claims histories—by showing that conditioning on claims experience leads to material refinements of the households' predicted claim rates. For instance, we find that (i) among households with downward revisions, their predicted claim rates decrease on average by 7 percent in auto collision, 13 percent in auto comprehensive, and 14 percent in home and (ii) among households with upward revisions, their predicted claim rates increase on average by 10 percent in auto collision, 23 percent in auto comprehensive, and 28 percent in home. Moreover, we demonstrate the incremental value of conditioning across lines of coverage (in addition to conditioning within lines of coverage) by showing that it not only leads to material incremental refinements of the households' predicted claim rates but also improves their accuracy.

Next, we show that experience rating—i.e., adjusting premiums to reflect households' a posteriori predicted claim rates—leads to material refinements of the households' premiums in home insurance.[2] For example, we find that (i) among households with downward adjustments, their premiums decrease on average by 11 percent and (ii) among households with upward adjustments, their premiums increase on average by 25 percent. As before, we demonstrate the incremental value of experience rating across lines of coverage (as opposed to experience rating only within lines of coverage) by showing that it leads to material incremental refinements of premiums.

Lastly, we investigate the extent to which households' demand for insurance, as captured by their deductible choices, is responsive to experience rating. To model households' deductible choices, we adopt the theoretical framework of Barseghyan et al. (2013) and consider two models of risky choice featured therein: (i) the standard expected utility model and (ii) a generalization of the expected utility model that allows for probability distortions. After calibrating the models using the parameter estimates reported by Barseghyan et al. (2013), we use both models to generate home deductible choices for the households in our data assuming first that premiums are not experience rated and then that they are experience rated. We find that (i) under the expected utility model, 14 percent percent of households would choose a different home deductible if their premiums were experience rated, and that (ii) under the probability distortion model, 11 percent of households would choose a different home deductible if their premiums were experience rated.

We close the paper by discussing legal restrictions on experience rating and what our findings imply about their economic consequences. In a nutshell, we argue that while our findings suggest that such restrictions lead to unpriced heterogeneity and in turn to distortions in households' demand for insurance, at the same time they suggest the magnitude of the demand distortions depends on the sources of insureds' aversion to risk. More specifically, we argue that our findings suggest that the size of the demand distortions arising from a failure to experience rate premiums depends on whether and to what ex-

[2] When we turn to experience rating, we restrict attention to home insurance out of necessity. Our rating model requires as an input the value of the insured property. We observe this value in the home policies but not in the auto policies.

tent insureds' risk aversion arises from diminishing marginal utility for wealth or from other sources such as subjective beliefs, probability weighting, loss aversion, or disappointment aversion.

The paper operates at the intersection two literatures. The first is the empirical literature on the regulation of insurance markets, and in particular the strand that seeks to quantify the economic effects of legal restrictions on risk classification by insurers. For example, Buchmueller and DiNardo (2002), Simon (2005), and Bundorf and Simon (2006) study the effects of community rating in U.S. health insurance markets. Finkelstein et al. (2009) study the effects of a ban on gender-based pricing in the U.K. annuity market. And Bundorf et al. (2012) and Geruso (2013) study the effects of uniform contribution requirements in the U.S. employer-provided health insurance market.[3]

The second is the literature on experience rating in insurance markets. For surveys, see Pinquet (2000, 2013) and Antonio and Valdez (2012).[4] Most closely related are the handful of papers on multidimensional experience rating, beginning with Jewell (1974). For example, Pinquet (1998) studies experience rating across claims at fault and not at fault. Desjardins et al. (2001) and Angers et al. (2006) study experience rating for fleets of vehicles. Englund et al. (2008) and Englund et al. (2009) study experience rating across various types of commercial coverage. Frees et al. (2010) study experience rating across multiple perils within home insurance. And Antonio et al. (2011) study experience rating across multiple auto insurance policies. To our knowledge, ours is the first paper to study experience rating across auto and home insurance and to assess, under different theories of risky choice, the demand effects of legal restrictions on within- and cross-coverage experience rating.

[3]Though it is not their focus, Einav et al. (2010) also consider the effects of legal restrictions on risk classification in the U.S. employer-provided health insurance market. There also is a rich theoretical literature on insurance regulation and the social welfare implications of legal restrictions on risk classification (e.g., Hoy 1982; Crocker and Snow 1986; Hoy 2006; Thomas 2008; Crocker and Snow 2011; Rothschild 2011)

[4]For textbook treatments, see Lemaire (1995), Bühlmann and Gisler (2005), and Denuit et al. (2007).

2 Description of the Data

The source of the data is a large U.S. property and casualty insurance company. The company offers several lines of insurance, including auto and home. The full data set includes annual information on more than $400,000$ households who held auto or home policies between 1998 and 2006. The data contain all the information in the company's records regarding the households and their policies (premiums, deductibles, etc.). In addition, the data record the number of claims that each household filed with the company under each of its policies during the period of observation.

We focus our attention on three lines of coverage: auto collision, auto comprehensive, and home all perils. Auto collision coverage pays for damage to the insured vehicle caused by a collision with another vehicle or object, without regard to fault. Auto comprehensive coverage pays for damage to the insured vehicle from all other causes (e.g., theft, fire, flood, windstorm, glass breakage, vandalism, hitting or being hit by an animal or by falling or flying objects), without regard to fault. Home all perils coverage pays for damage to the insured home from all causes (e.g., fire, windstorm, hail, tornadoes, vandalism, or smoke damage), except those that are specifically excluded (e.g., flood, earthquake, or war). For simplicity, we often refer to home all perils merely as home.

In most of the analysis, we consider an unbalanced panel of $62,425$ households who held all three coverages (auto collision, auto comprehensive, and home) in one or more years between 1998 and 2006. In all, this tricoverage sample comprises $294,917$ household-years. Descriptive statistics are set forth in the Appendix.

Table 1 summarizes the claims, premiums, and deductibles in the tricoverage sample. Additional details are set forth in the Appendix. The mean number of claims per household-year is 0.107 in auto collision, 0.032 in auto comprehensive, and 0.079 in home. On average, households paid annual premiums of $200 in auto collision, $127 in auto comprehensive, and $548 in home. The mean deductibles per household-year are $396, $273, and $350 in auto collision, auto comprehensive, and home, respectively. The modal deductibles are \$500 in auto collision, \$200 in auto comprehensive, and \$250 in home.

3 Model and Estimation Strategy

A standard regression model for longitudinal univariate count data is the Poisson random effects model. We extend this model to multivariate count data—here, claim counts under three types of insurance coverage—by allowing for correlated random effects.[5]

Let y_{itk} denote the number of claims for household i in year t under coverage k, where $i = 1, ..., N$, $t = 1, ..., T_i$, and $k \in \{c, m, h\}$.[6] Similarly, let denote a vector of observables (plus a constant) for household i in year t under coverage k.[7] Let λ_{itk} denote household i's baseline claim rate in year t under coverage k, and let ϵ_{ik} denote a time-constant random effect for household i under coverage k. Both λ_{itk} and ϵ_{ik} are unobserved.

We assume

$$y_{itk}|\mathbf{x}_{itk} \sim Poisson\left(\lambda_{itk}\epsilon_{ik}\right),$$

where

$$\lambda_{itk} = \exp\left(\mathbf{x}'_{itk}\boldsymbol{\beta}_k\right)$$

and $\boldsymbol{\epsilon}_i \equiv [\epsilon_{ic}\ \epsilon_{im}\ \epsilon_{ih}]'$ is iid with $E(\boldsymbol{\epsilon}_i) = [1\ 1\ 1]'$ and $V(\boldsymbol{\epsilon}_i) = \boldsymbol{\Sigma}$.

[5] By allowing for random effects, our model accounts for overdispersion, including due to excess zeros, in a similar way as the (pooled) negative binomial model (see, e.g., Wooldridge 2002, ch. 19). An alternative approach would be a zero-inflated model. However, Vuong (1989) and likelihood ratio tests select the negative binomial model over the zero-inflated model, suggesting that adjustment for excess zeros is not necessary once we allow for random effects.

[6] In the set of coverages, c denotes auto collision, m denotes auto comprehensive, and h denotes home.

[7] The variables that comprise $\mathbf{x}_{itk}$ are listed in Tables A5 (auto) and A6 (home) in the Appendix. All of the variables are observed by the company and used by the company to underwrite and rate its policies. In auto, they include the age, gender, and insurance score (which is based on information contained in credit reports) of the primary driver, the age and gender of each additional driver, and the age, use, location, and safety features of each vehicle. In home, they include the insurance score of the primary owner, the age, value, use, location, type of construction, and safety features of the dwelling, whether the dwelling is owner occupied, and the number of families that occupy the dwelling.

The parameters to be estimated are

$$\boldsymbol{\beta} \equiv \begin{bmatrix} \boldsymbol{\beta}_c \\ \boldsymbol{\beta}_m \\ \boldsymbol{\beta}_h \end{bmatrix} \quad \text{and} \quad \boldsymbol{\Sigma} \equiv \begin{bmatrix} \sigma_c^2 & \rho_{mc}\sigma_m\sigma_c & \rho_{hc}\sigma_h\sigma_c \\ \rho_{cm}\sigma_c\sigma_m & \sigma_m^2 & \rho_{hm}\sigma_h\sigma_m \\ \rho_{ch}\sigma_c\sigma_h & \rho_{mh}\sigma_m\sigma_h & \sigma_h^2 \end{bmatrix}.$$

Of principal interest is the variance-covariance matrix, $\boldsymbol{\Sigma}$, which captures both the within-coverage variance of unobserved heterogeneity, $\boldsymbol{\sigma^2} \equiv (\sigma_c^2, \sigma_m^2, \sigma_h^2)$, and its cross-coverage correlation structure, $\boldsymbol{\rho} \equiv (\rho_{cm}, \rho_{ch}, \rho_{mh})$.

The likelihood function may be written as

$$\mathfrak{L}_i = \int_{\epsilon_{ih}} \int_{\epsilon_{im}} \int_{\epsilon_{ic}} \left\{ \prod_k \prod_t \exp\left(-\lambda_{itk}\epsilon_{ik}\right) \frac{\left(\lambda_{itk}\epsilon_{ik}\right)^{y_{itk}}}{y_{itk}!} \right\} f(\epsilon_{ic}, \epsilon_{im}, \epsilon_{ih}) d\epsilon_{ic} d\epsilon_{im} d\epsilon_{ih},$$

where $f(\epsilon_{ic}, \epsilon_{im}, \epsilon_{ih})$ is the trivariate density of ϵ_i. A standard parametric approach is to specify f and estimate the model by maximum likelihood. Typical specifications of f include the lognormal distribution and the gamma distribution (in which case $\mathfrak{L}_i$ reduces to the product of negative binomial densities). In our case, however, the standard approach is computationally intractable. The likelihood function not only involves a multidimensional integral, but, depending on f, it also may not have a closed-form expression.

We adopt a semiparametric, moments-based approach, which provides a computationally tractable method for consistent estimation of $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ for all possible densities f. Under this approach, estimation is via generalized estimating equations (GEE) based on marginal moments.[8] Given the assumptions of our model, we can derive the first and second marginal moments and use them to construct estimating equations for $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$. More specifically, we use the first marginal moment to define a quasi-score equation, where the associated estimating equation for $\boldsymbol{\beta}$ is based on a weighted least squares estimator with the weight matrix defined by the covariance structure derived from the second marginal moment. The estimating equation for $\boldsymbol{\Sigma}$ is based

[8] GEE were introduced by Liang and Zeger and co-authors in the 1980s (see, e.g., Liang and Zeger 1986; Zeger and Liang 1986; Zeger et al. 1988). For a textbook treatment of GEE, see, e.g., Ziegler (2011).

on the relation between the empirical variance estimate and the model defined covariance structure. The two estimating equations are solved iteratively to obtain $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\Sigma}}$. For further details about the estimation approach, see the Appendix. See also Morris (2012) and Pinquet (2013).[9]

4 Estimation Results

4.1 Regression Estimates

Table 2 presents the estimates of the association parameters, $\boldsymbol{\sigma^2}$ and $\boldsymbol{\rho}$, implied by $\widehat{\boldsymbol{\Sigma}}$. The estimates reveal that the variance of unobserved heterogeneity is lowest in auto collision ($\widehat{\sigma}_c^2 = 0.11$) and is roughly four times higher in auto comprehensive ($\widehat{\sigma}_m^2 = 0.40$) and home ($\widehat{\sigma}_h^2 = 0.41$). The estimates also reveal that unobserved heterogeneity is correlated across coverages—each pairwise correlation is positive and statistically significant—suggesting that there is a domain-general component to risk type. Perhaps not surprisingly, the strongest correlation is between auto collision and auto comprehensive ($\widehat{\rho}_{cm} = 0.66$). More surprising, however, is the fairly strong correlation between auto comprehensive and home ($\widehat{\rho}_{mh} = 0.56$). After all, one might conjecture that claim risk in auto comprehensive and home reflect force majeure risk more than household risk. The weakest correlation is between auto collision and home ($\widehat{\rho}_{ch} = 0.29$). Even this correlation, however, is economically significant, as we demonstrate in Section 5.

The estimates of the regression parameters, $\boldsymbol{\beta}$, are reported in the Appendix. Because $\boldsymbol{\beta}$ is not the object of primary interest, we relegate our comments about the regression parameter estimates to the Appendix as well.

[9] This approach is an extension of quasi-generalized pseudo maximum likelihood (QGPML) estimators developed by Gouriéroux et al. (1984a,b) and the extended GEE approach developed by Prentice (1988). The QGPML method can be characterized as first order GEE with a specific association structure. Prentice introduced an extension of first order GEE that utilizes a second set of estimating equations to jointly estimate the association parameters. QGPML can be embedded in the GEE framework resulting in commonly studied consistency and asymptotic results for simultaneous inference on both the regression parameters and the association parameters.

4.2 Alternative Samples

As checks of the sensitivity of the association parameter estimates, we re-estimate the model on two alternative samples of the data: (A) a balanced panel of $8,731$ households ($78,579$ household-years) who held all three coverages (auto collision, auto comprehensive, and home); and (B) an unbalanced panel of $203,731$ households ($1,019,170$ household-years) who held both auto coverages (collision and comprehensive). The association parameter estimates for both alternative samples are reported in the Appendix.[10] They are largely consistent with the estimates for the tricoverage sample.

4.3 Moral Hazard

Our approach implicitly assumes that a household's claim risk is not a function of its choice of deductible. That is, we assume households do not suffer from moral hazard. In particular, we assume there is neither ex ante moral hazard (deductible choice does not influence the frequency of claimable events) nor ex post moral hazard (deductible choice does not influence the decision to file a claim). The empirical evidence on moral hazard in auto insurance markets is mixed (Cohen and Siegelman 2010), and we are not aware of any empirical evidence on moral hazard in home insurance markets. Because deductibles are small relative to the overall level of coverage, it seems reasonable to assume there is no ex ante moral hazard. However, because the damage from a claimable event may occasionally be less than the chosen deductible (at least for "high deductible" households), it may be less reasonable to assume there is no ex post moral hazard. As a check of the sensitivity of the association parameter estimates to our assumption on moral hazard, we re-estimate the model separately for "low deductible" and "high deductible" households. We

[10] To ease the computational burden of the re-estimation, we obtain estimates of the regression parameters from a generalized linear model (GLM) assuming the random effects follow a lognormal distribution. In the tricoverage sample, the semiparametric and GLM estimates for $\boldsymbol{\beta}$ are nearly identical ($R^2 = 0.9998$). Thus, we are confident that using the GLM estimates for $\boldsymbol{\beta}$ does not corrupt the semiparametric estimates of the association parameters in the re-estimation.

define a household as a "low deductible" household if none of its deductibles is greater than \$250. Conversely, we define a household as a "high deductible" household if at least one of its deductibles is greater than \$250. Table 3 reports the association parameter estimates for low and high deductible households.[11] They are largely consistent with each other and with the estimates for the tricoverage sample, suggesting that moral hazard is not an issue.

5 Information Value of Claims Experience

In this section, we demonstrate the value of the information contained in the estimated variance-covariance matrix $\widehat{\mathbf{\Sigma}}$—and, by implication, the information value of the households' claims histories—by showing that conditioning on claims experience leads to material refinements of the predicted claim rates in the tricoverage sample. We also demonstrate the incremental value of utilizing the information on the cross-coverage correlation structure of unobserved heterogeneity ($\boldsymbol{\rho}$), as opposed to utilizing only the information on the within-coverage variance of unobserved heterogeneity ($\boldsymbol{\sigma}^2$), by showing that conditioning across lines of coverage (in addition to conditioning within lines of coverage) leads to material incremental refinements of the predicted claim rates and also improves their accuracy.

Throughout this section and beyond, we distinguish among three types of predicted claim rates. The first are *prior claim rates*, $\widehat{\lambda}_{itk} \equiv \exp(\mathbf{x}'_{itk}\widehat{\boldsymbol{\beta}}_k)$. These are a priori predicted claim rates based on ex ante observables. The second are *univariate posterior claim rates*, $\widehat{\vartheta}_{itk} \equiv \widehat{\lambda}_{itk}E^{UV}(\epsilon_{ik}|\mathbf{y}_{ik})$ for each coverage $k = c, m, h$, where $\mathbf{y}_{ik} \equiv (y_{i1k}, ..., y_{iT_ik})$ and $E^{UV}(\epsilon_{ik}|\mathbf{y}_{ik})$ is calculated assuming $\epsilon_{ik} \overset{iid}{\sim}$ *lognormal* with $E(\epsilon_{ik}) = 1$ and $V(\epsilon_{ik}) = \widehat{\sigma}^2_k$. These are a posteriori predicted claim rates conditional on within-coverage ex post claims experience. The third are *multivariate posterior claim rates*, $\widehat{\theta}_{itk} \equiv \widehat{\lambda}_{itk}E^{MV}(\epsilon_{ik}|\mathbf{y}_i)$, where $\mathbf{y}_i \equiv (\mathbf{y}_{ic}, \mathbf{y}_{im}, \mathbf{y}_{ih})$ and $E^{MV}(\epsilon_{ik}|\mathbf{y}_i)$ is calculated assuming $\boldsymbol{\epsilon}_i \equiv [\epsilon_{ic}\ \epsilon_{im}\ \epsilon_{ih}]' \overset{iid}{\sim}$ *lognormal* with $E(\boldsymbol{\epsilon}_i) = [1\ 1\ 1]'$ and $V(\boldsymbol{\epsilon}_i) = \widehat{\mathbf{\Sigma}}$.

[11] As before, the re-estimations use GLM estimates of the regression parameters assuming the random effects follow a lognormal distribution.

These are a posteriori predicted claim rates conditional on ex post claims experience both within and across coverages.[12]

5.1 Updating

In order to demonstrate that conditioning on claims experience leads to material refinements of the predicted claim rates in the tricoverage sample, we compare the empirical distribution of the prior claim rates, $\widehat{\lambda}_{itk}$, with that of the multivariate posterior claim rates, $\widehat{\theta}_{itk}$. Figure 1 plots, for each coverage $k = c, m, h$, the kernel density of $\eta_{itk} \equiv (\widehat{\theta}_{itk} - \widehat{\lambda}_{itk})/\widehat{\lambda}_{itk}$. Further details are set forth in Table 4. For households with negative values of η_{itk}, the mean value of η_{itk} is -7 percent in auto collision, -13 percent in auto comprehensive, and -14 percent in home. For a quarter of these households, η_{itk} is less than -9 percent in auto collision, -19 percent in auto comprehensive, and -20 percent in home. For a tenth of these households, η_{itk} is less than -12 percent in auto collision and -24 percent in both auto comprehensive and home. The numbers are even more striking for households with positive values of η_{itk}.[13] For these households, the mean value of η_{itk} is $+10$ percent in auto collision, $+23$ percent in auto comprehensive, and $+28$ percent in home. For a quarter of these households, η_{itk} exceeds $+14$ percent in auto collision, $+31$ percent in auto comprehensive, and $+37$ percent in home. For a tenth of these households, η_{itk} exceeds $+23$ percent in auto collision, $+53$ percent in auto comprehensive, and $+65$ percent in home. The numbers are remarkably similar for households with low, medium, and high prior claim rates,[14] suggesting that the value of the information in $\widehat{\mathbf{\Sigma}}$ is robust to differences in baseline claim risk.

To show the incremental value of conditioning across lines of coverage, we compare the empirical distribution of the multivariate posterior claim rates,

[12] The derivations of $E^{UV}(\epsilon_{ik}|\mathbf{y}_{ik})$ and $E^{MV}(\epsilon_{ik}|\mathbf{y}_{i})$ are set forth in the Appendix.

[13] This is readily explained by the fact that the prior claim rates are relatively small (98.7 percent of the prior claim rates are less than 0.2), which implies that experiencing a claim generally is more informative than not experiencing a claim.

[14] A prior claim rate is "low" if it is in the bottom quartile and "high" if it is in the top quartile. It is "medium" otherwise. In the tricoverage sample, the respective low and high cutoffs are 0.078 and 0.127 in auto collision, 0.016 and 0.044 in auto comprehensive, and 0.054 and 0.096 in home.

$\widehat{\theta}_{itk}$, with that of the univariate posterior claim rates, $\widehat{\vartheta}_{itk}$. Figure 2 plots, for each coverage k, the kernel density of $\zeta_{itk} \equiv (\widehat{\theta}_{itk} - \widehat{\vartheta}_{itk})/\widehat{\vartheta}_{itk}$. Further details are set forth in Table 5. For households with negative values of ζ_{itk}, the mean value of ζ_{itk} is -3 percent in auto collision, -10 percent in auto comprehensive, and -4 percent in home, and for a tenth of these households ζ_{itk} is less than -6 percent in auto collision, -17 percent in auto comprehensive, and -8 percent in home. Again, the numbers are more striking for households with positive values of η_{itk}. For these households, the mean value of ζ_{itk} is $+7$ percent in auto collision, $+16$ percent in auto comprehensive, and $+9$ percent in home, and for a tenth of these households ζ_{itk} exceeds $+15$ percent in auto collision, $+36$ percent in auto comprehensive, and $+21$ percent in home. As before, the numbers are very similar for households with low, medium, and high prior claim rates, suggesting that the incremental value of the information in $\widehat{\boldsymbol{\rho}}$ is robust to differences in baseline claim risk.

5.2 Accuracy

In this section, we show that conditioning across lines of coverage (in addition to conditioning within lines of coverage) improves the accuracy of the predicted claim rates. To do so, we move from the actual data to simulated data. The key virtue of the simulated data is that we can observe the "true" baseline claim rates, λ_{itk}, and "true" variance-covariance matrix, $\boldsymbol{\Sigma}$, neither of which is observable in the actual data. This allows us to make statements about the relative accuracy of the univariate and multivariate posterior claim rates.

We consider 18 cases. In each case, there are $10,000$ identical households. The cases differ on three variables: the households' baseline claim rates, $\boldsymbol{\lambda} \equiv (\lambda_c, \lambda_m, \lambda_h)$; the time horizon, T; and the variance-covariance matrix, $\boldsymbol{\Sigma}$. We consider three levels of baseline claim rates: (1) "average" baseline claim rates, which roughly correspond to the mean prior claim rates in the tricoverage sample: $\boldsymbol{\lambda} = (0.100, 0.030, 0.070)$; (2) "low" baseline claim rates, which correspond to the 25th percentiles: $\boldsymbol{\lambda} = (0.078, 0.016, 0.054)$; and (3) "high" baseline claim rates, which correspond to the 75th percentiles:

$\boldsymbol{\lambda} = (0.127, 0.044, 0.096)$. We also consider three time horizons: $T = 1$; $T = 5$; and $T = 10$. Finally, we consider two specifications for the variance-covariance matrix.[15] In specification A, we set $\boldsymbol{\Sigma} = \widehat{\boldsymbol{\Sigma}}$ to match the estimates from the data. In specification B, we fix the within-coverage variances such that each equals 0.250 and we adjust the cross-coverage covariances such that the cross-coverage correlations still match the estimates from the data (i.e., $\boldsymbol{\rho} = \widehat{\boldsymbol{\rho}}$). Hence, in specification A $\boldsymbol{\sigma^2} = (0.107, 0.399, 0.405)$ and $\boldsymbol{\rho} = (0.663, 0.293, 0.559)$, which yields

$$\boldsymbol{\Sigma} = \begin{bmatrix} 0.107 & 0.137 & 0.061 \\ 0.137 & 0.399 & 0.225 \\ 0.061 & 0.225 & 0.405 \end{bmatrix} \quad \text{(specification A)},$$

and in specification B $\boldsymbol{\sigma^2} = (0.250, 0.250, 0.250)$ and $\boldsymbol{\rho} = (0.663, 0.293, 0.559)$, which yields

$$\boldsymbol{\Sigma} = \begin{bmatrix} 0.250 & 0.166 & 0.073 \\ 0.166 & 0.250 & 0.140 \\ 0.073 & 0.140 & 0.250 \end{bmatrix} \quad \text{(specification B)}.$$

The virtue of moving from specification A to specification B is that it changes the within-coverage variances (and fixes them at a "middle" level) while holding constant the cross-coverage correlation structure. This serves to isolate the independent value of the information in $\boldsymbol{\rho}$, for we can assess the extent to which the results are—or are not—being driven by $\boldsymbol{\sigma^2}$.

We perform 500 iterations of each case. In each iteration $j = 1, ..., 500$: (i) we simulate the (time-constant) random effect ϵ_{ij} for each household—i.e., for each household $i = 1, ..., 10,000$, we draw $\epsilon_{ij} \equiv [\epsilon_{ijc}\ \epsilon_{ijm}\ \epsilon_{ijh}]'$ from a lognormal distribution with $E(\boldsymbol{\epsilon}_{ij}) = [1\ 1\ 1]'$ and $V(\boldsymbol{\epsilon}_{ij}) = \boldsymbol{\Sigma}$; (ii) we simulate the claims experience $\mathbf{y}_{ij}$ of each household—i.e., for each household $i = 1, ..., 10,000$ and each year $t = 1, ..., T$, we draw y_{itk} from $Poisson\,(\lambda_k \epsilon_{ijk})$ for each coverage $k = c, m, h$; (iii) we estimate the model on the simulated data—i.e., we obtain

[15] Of course, we fix $E([\epsilon_{ic}\ \epsilon_{im}\ \epsilon_{ih}]') = [1\ 1\ 1]'$ in all cases.

$\widehat{\boldsymbol{\Sigma}}_j$; and (iv) we calculate the univariate and multivariate posterior claim rates for each household—i.e., for each household $i = 1, ..., 10,000$, we calculate $\widehat{\vartheta}_{ijk} \equiv \lambda_k E^{UV}(\epsilon_{ijk}|\mathbf{y}_{ij})$ and $\widehat{\theta}_{ijk} \equiv \lambda_k E^{MV}(\epsilon_{ijk}|\mathbf{y}_{ij})$ for each coverage $k = c, m, h$.[16] We then compute, for each coverage $k = c, m, h$,

$$MSE_k^{UV} \equiv \frac{1}{10,000}\frac{1}{500}\sum_{i=1}^{10,000}\sum_{j=1}^{500}\left(\widehat{\vartheta}_{ijk} - \lambda_k\epsilon_{ijk}\right)^2$$

and

$$MSE_k^{MV} \equiv \frac{1}{10,000}\frac{1}{500}\sum_{i=1}^{10,000}\sum_{j=1}^{500}\left(\widehat{\theta}_{ijk} - \lambda_k\epsilon_{ijk}\right)^2,$$

as well as

$$MSE_{ik}^{UV} \equiv \frac{1}{500}\sum_{j=1}^{500}\left(\widehat{\vartheta}_{ijk} - \lambda_k\epsilon_{ijk}\right)^2$$

and

$$MSE_{ik}^{MV} \equiv \frac{1}{500}\sum_{j=1}^{500}\left(\widehat{\theta}_{ijk} - \lambda_k\epsilon_{ijk}\right)^2$$

for each household $i = 1, ..., 10,000$ and

$$MSE_{jk}^{UV} \equiv \frac{1}{10,000}\sum_{i=1}^{10,000}\left(\widehat{\vartheta}_{ijk} - \lambda_k\epsilon_{ijk}\right)^2$$

and

$$MSE_{jk}^{MV} \equiv \frac{1}{10,000}\sum_{i=1}^{10,000}\left(\widehat{\theta}_{ijk} - \lambda_k\epsilon_{ijk}\right)^2$$

for each iteration $j = 1, ..., 500$.

Table 6 summarizes the results. There are two main takeaways. The first main takeaway is that, in every case, $MSE_k^{MV} < MSE_k^{UV}$ for each coverage. In cases with a one-year time horizon, the multivariate posterior claim rates

[16] Strickly speaking, we should use the average claim count $\overline{y}_k$ instead of λ_k in calculating $\widehat{\vartheta}_{ijk}$ and $\widehat{\theta}_{ijk}$, to make them directly comparable to $\widehat{\vartheta}_{itk}$ and $\widehat{\theta}_{itk}$. However, we use λ_k to abstract from the statistical uncertainty in estimating prior claim rates and focus on the value of utilizing the information in $\widehat{\boldsymbol{\Sigma}}$.

are only slightly more accurate than the univariate posterior claim rates. In these cases, the absolute errors of the multivariate posterior claim rates are 0.1 percent to 1.0 percent less than the absolute errors of the univariate posterior claim rates. However, the accuracy advantage of the multivariate posterior claim rates is more pronounced in cases with higher baseline claim rates and longer time horizons. Indeed, in cases with high baseline claim rates and a ten-year time horizon, the absolute errors of the multivariate posterior claim rates are 2.0 percent to 7.6 percent less than the absolute errors of the univariate posterior claim rates. This increase in accuracy is due to the increase in the percentage of households for whom $MSE_{ik}^{MV} < MSE_{ik}^{UV}$, which increases from roughly 70 to 80 percent in cases with low claim rates and a one-year time horizon to more than 95 percent in cases with high claim rates and a ten-year time horizon. By contrast, $MSE_{jk}^{MV} < MSE_{jk}^{UV}$ in virtually every iteration of every case. The second main takeaway is that the accuracy advantage of the multivariate posterior claim rates is essentially the same for both specifications of $\mathbf{\Sigma}$. This suggests that, in terms of improving the accuracy of the predicted claim rates, the value of the information on the cross-coverage correlation structure of unobserved heterogeneity does not depend on the structure of the within-coverage variance of unobserved heterogeneity.

6 Experience Rating and Deductible Choices

The previous section demonstrates the information value of households' claims histories by showing that conditioning on claims experience leads to material refinements of the predicted claim rates and improves their accuracy. In this section, we show that experience rating—i.e., adjusting premiums to reflect households' posterior claim rates—leads to material refinements of the households' premiums.[17] We then investigate the extent to which households' demand for insurance—as captured by their deductible choices—would respond

[17]Under experience rating, an insured's premium is adjusted based on his or her actual loss experience. Experience rating is not to be confused with classification rating, under which an insured's premium is based on the collective loss experience of all insureds in the insured's risk class. See, e.g., N.Y. Ins. Dep't OGC Op. No. 08-07-16 (2008).

to experience rating.

6.1 Experience Rating

To show that experience rating leads to material refinements in premiums, we conduct a two-step exercise. First, we develop a simple model of insurance premiums and estimate the risk elasticity of price—i.e., the elasticity of premiums with respect to predicted claim rates. Second, we use the elasticity estimates to assess how premiums would respond to moving from prior to posterior claim rates.

6.1.1 Model

We take a textbook approach to modeling insurance premiums (see, e.g., Schlesinger 2000; Eeckhoudt et al. 2005). To begin, we assume that (i) there is at most one claim during the policy period[18] and (ii) the base price $\overline{p}_i$ for a policy i is proportional to its actuarial value:

$$\overline{p}_i = \kappa_i \widehat{\mu}_i [E(L_i) - \overline{d}],$$

where $\overline{d}$ is the base deductible, κ_i is the loading factor (i.e., the loading for profits), $\widehat{\mu}_i$ is the predicted probability of a claim, E is the expectation operator, and L_i is the random loss in the event of a claim.

The base deductible is simply a benchmark amount that the company uses for purposes of determining the base price. The base deductible is \$200 in auto and \$250 in home. The company then uses the base price to generate a menu of premium-deductible pairs, $\mathcal{M} = \{(p(d), d) : d \in \mathcal{D}\}$, where $p_i(d)$ is the premium for coverage with deductible d and $\mathcal{D}$ is the set of deductible options. For every deductible d, $p_i(d)$ is simply an affine transformation of the base price $\overline{p}_i$. In particular, $p_i(d) = g(d)\overline{p}_i + c_i$, where $g(d) > 0$ and $c_i > 0$.[19]

[18] Because 98.7 percent of the prior claim rates in the tricoverage sample are less than 0.2, the likelihood of two or more claims is very small.

[19] Naturally, $g(\overline{d}) = 1$ and $g'(d) < 0$. The multiplicative factors $\{g(d) : d \in \mathcal{D}\}$ are known in the industry as the deductible factors and c_i is a small markup known as the expense fee.

We assume the loading factor varies from policy to policy: $\kappa_i = \mathbf{z}_i'\boldsymbol{\alpha}$, where $\mathbf{z}_i$ is a vector of policy-specific variables (plus a constant) and $\boldsymbol{\alpha}$ is a vector of coefficients.[20] We also assume that claims follow a Poisson process with intensity λ_i, and we assume initially that $\lambda_i = \widehat{\lambda}_i$. Together with the assumption that there is at most one claim during the policy period, this implies $\widehat{\mu}_i = 1 - \exp(-\widehat{\lambda}_i)$.[21] We further assume that L_i follows a Beta distribution with shape parameters a and b on the support set $[l, v_i]$, where l and v_i are the minimum and maximum possible loss in the event of a claim, respectively. This implies $E(L_i) = (v_i - l)[a/(a+b)] + l$. Finally, we assume $l = \overline{d}$. This ensures $E(L_i) > \overline{d}$ and thus $\overline{p}_i > 0$.

Given these assumptions, we can re-write the model as follows:

$$\overline{p}_i = \mathbf{x}_i'\boldsymbol{\gamma},$$

where $\mathbf{x}_i' = \left[\widehat{\mu}_i(v_i - \overline{d})\right]\mathbf{z}_i'$ and $\boldsymbol{\gamma} = [a/(a+b)]\boldsymbol{\alpha}$.

6.1.2 Estimation and Results

In the estimation, we restrict attention to home policies in the tricoverage sample. This is because we observe the insured value of the home, which we set equal to v_i. (In the auto policies, we do not observe the insured value of the vehicle.) We also observe the base deductible, $\overline{d} = \$250$. In addition, we exclude renewal policies. This avoids any concerns about price inertia/rigidities (i.e., concerns that premiums may not be adjusted at the time of renewal to reflect changes in $\mathbf{z}_i$). Consequently, our estimation sample comprises a cross section of 62,425 home policies. We refer to this sample as the home sample.

The least squares estimates of $\boldsymbol{\gamma}$ are reported in the Appendix. Table 7 summarizes the fitted values of $\overline{p}_i$, which we denote by $\widehat{p}_i$, and the implied risk elasticities of price, $\xi_i \equiv (\partial\widehat{p}_i(d^*)/\partial\widehat{\mu}_i)[\widehat{\mu}_i/\widehat{p}_i(d^*)]$, where $\widehat{p}_i(d) = g(d)\widehat{p}_i + c_i$

[20] The variables that comprise $\mathbf{z}_i$ are listed in Table A8 in the Appendix. They are a subset of the variables that comprise $\mathbf{x}_{itk}$ in Section 3.

[21] The Poisson probability mass function is $f(y, \lambda) = \exp(-\lambda)\lambda^y/y!$ for $y = 0, 1, 2, ...$ and $\lambda \geq 0$. Thus, if the number of claims y follow a Poisson distribution with arrival rate λ, then the probability of experiencing at least one claim is $1 - \exp(-\lambda)$.

denotes the fitted premium for coverage with deductible d and d^* denotes the observed deductible choice. (In addition, Figure 3 compares the kernel density of $\widehat{p}_i$ with that of $\overline{p}_i$.) There are two main takeaways from Table 7. First, despite its simplicity, the model fits the data reasonably well: the adjusted R^2 is 0.83 and the distributions of $\widehat{p}_i$ and $\overline{p}_i$ are quite similar. Second, the estimates imply sizeable risk elasticities of price: the mean implied value of ξ_i is 0.82 and the 25th, 50th, and 75th percentiles are 0.77, 0.84, and 1.00, respectively. This suggests that premiums are amply responsive to changes in predicted claim rates, which in turn suggests that experience rating would lead to material changes in premiums.

To get a more precise idea of how premiums would respond to experience rating, we examine how the distribution of $\widehat{p}_i$ changes when premiums are calculated using *multivariate posterior claim probabilities*, $\widehat{\mu}_i^{MV} = 1 - \exp(-\widehat{\theta}_i)$, i.e., when premiums are experience rated within and across coverages. Figure 4 plots the kernel density of $\Gamma_i \equiv (\widehat{p}_i^{MV} - \widehat{p}_i)/\widehat{p}_i$, where $\widehat{p}_i^{MV}$ denotes premiums calculated using multivariate posterior claim probabilities. Further details are set forth in Table 8. The main takeaway is that premiums are highly responsive to experience rating. For policies with negative values of Γ_i, the mean value of Γ_i is -11 percent. For a quarter of these policies Γ_i is less than -16 percent, and for a tenth Γ_i is less than -21 percent. For policies with positive values of Γ_i, the mean value of Γ_i is $+25$ percent. For a quarter of these policies Γ_i is more than $+33$ percent, and for a tenth Γ_i is more than $+58$ percent. There is some variation in the numbers for policies with low, medium, and high prior claim probabilities;[22] however, the breakdowns clearly evince that the main takeaway is robust to differences in baseline claim risk.

To get an idea of the incremental effect of experience rating across coverages, as opposed to experience rating only within coverages, we compare the distribution of $\widehat{p}_i^{MV}$ with that of $\widehat{p}_i^{UV}$, where $\widehat{p}_i^{UV}$ denotes premiums calculated

[22] For instance, for policies with negative values of Γ_i, the mean value of Γ_i for policies with low, medium, and high prior claim probabilities is -7 percent, -12 percent, and -13 percent, respectively. This variation reflects the fact that the information value of not experiencing a claim is greatest for policies with high prior claim probabilities and lowest for policies with low prior claim probabilities.

using *univariate posterior claim probabilities*, $\widehat{\mu}_i^{UV} = 1 - \exp(-\widehat{\vartheta}_i)$. Figure 4 plots the kernel density of $\Delta_i \equiv (\widehat{p}_i^{MV} - \widehat{p}_i^{UV})/\widehat{p}_i^{UV}$. Further details are set forth in Table 8. The main takeaway here is that cross-coverage experience rating has a material incremental effect in terms of further refining premiums that already have been experience rated within coverages. For policies with negative values of Δ_i, the mean value of Δ_i is -3 percent. For a quarter of these policies Δ_i is less than -5 percent, and for a tenth Δ_i is less than -7 percent. For policies with positive values of Δ_i, the mean value of Δ_i is $+9$ percent. For a quarter of these policies Δ_i is more than $+13$ percent, and for a tenth Δ_i is more than $+19$ percent. Again, although there is some variation in the numbers for policies with low, medium, and high prior claim probabilities, the main takeaway is robust to differences in baseline claim risk.

6.2 Deductible Choices

We now investigate the extent to which households' deductible choices would respond to experience rating. As before, we proceed in two steps. First, we adopt the theoretical framework of Barseghyan et al. (2013) [hereafter, BMOT] and consider two models of risky choice featured therein: (i) the standard expected utility model and (ii) a generalization of the expected utility model that allows for probability distortions. Second, we investigate how deductible choices would change under both models when we move from calculating premiums using prior claim probabilities to calculating premiums using univariate and multivariate posterior claim probabilities.

6.2.1 Two Models of Deductible Choice

A household i faces a menu of premium-deductible pairs $\{(p_i(d), d) : d \in \mathcal{D}\}$, where $p_i(d)$ is the household's premium for coverage with deductible d and $\mathcal{D}$ is the set of deductible options. The household experiences at most one claim during the policy period, and it believes the probability of experiencing a claim is μ_i. In the event of a claim, the loss exceeds the maximum deductible option and payment of the deductible is the only cost associated with the claim. Under these assumptions, the household's choice of deductible involves a choice among lotteries of the form

$$\mathcal{L}_i(d) \equiv (-p_i(d), 1-\mu_i; -p_i(d)-d, \mu_i)\,.$$

Under the standard expected utility model, the utility of lottery $\mathcal{L}_i(d)$ is given by

$$U_i(\mathcal{L}_i(d)) = (1-\mu_i)\,u_i\,(w_i - p_i(d)) + \mu_i u_i\,(w_i - p_i(d) - d)\,,$$

where $u_i\,(\cdot)$ is the household's Bernoulli utility function and w_i is its wealth.

BMOT also consider a generalization of the expected utility model that allows for probability distortions. Under this model, the utility of lottery $\mathcal{L}_i(d)$ is given by

$$U_i(\mathcal{L}_i(d)) = (1-\Omega_i(\mu_i))\,u_i\,(w_i - p_i(d)) + \Omega_i(\mu_i)u_i\,(w_i - p_i(d) - d)\,, \qquad (1)$$

where $\Omega_i\,(\cdot)$ is the household's probability distortion function. Given our setting, this model is quite general in the sense that it includes several others as special cases, including models of subjective beliefs, rank-dependent probability weighting (Quiggin 1982), loss aversion (Kőszegi and Rabin 2006, 2007), and disappointment aversion (Gul 1991). For details, see BMOT.

6.2.2 Estimation

BMOT estimate both models using a cross section of 4,170 households in the tricoverage sample. They arrive at their estimation sample by imposing two restrictions. First, they restrict attention to households who hold both auto and home policies and who first purchased their auto and home policies from the company in the same year, in either 2005 or 2006. This is meant to avoid temporal issues, such as changes in household characteristics and in the economic environment. Second, they consider only the initial deductible choices of each household. This is meant to increase confidence that they are working with active choices; one might be concerned that some households renew their policies without actively reassessing their deductible choices.

Although BMOT take several approaches, we focus on their benchmark case, in which they assume homogeneous preferences—i.e., every household has the same utility function $u(\cdot)$ and probability distortion function $\Omega(\mu_i)$. For $u(\cdot)$, they consider a second-order Taylor expansion. Also, because $u(\cdot)$ is unique only up to an affine transformation, they normalize the scale of utility by dividing $u'(\cdot)$. With this specification, equation (1) becomes

$$U(\mathcal{L}_i(d)) = -\left[p_i(d) + \Omega(\mu_i)d\right] - \frac{r}{2}\left[(1-\Omega(\mu_i))\left(p_i(d)\right)^2 + \Omega(\mu_i)\left(p_i(d)+d\right)^2\right], \quad (2)$$

where r is the coefficient of absolute risk aversion. In the expected utility model, BMOT set $\Omega(\mu_i) = \mu_i$. In the probability distortion model, they assume $\Omega(\mu_i) = \exp(\delta_0 + \delta_1\mu_i + \delta_2\mu_i^2)$. In both models, they assume $\mu_i = 1 - \exp(-\widehat{\vartheta}_i)$.

To account for observationally equivalent households choosing different deductibles, BMOT assume random utility with additive choice noise. Specifically, they assume that the utility from deductible $d \in \mathcal{D}$ is given by

$$\mathcal{U}(d) \equiv U(\mathcal{L}_i(d)) + \varepsilon_{i,d},$$

where $\varepsilon_{i,d}$ follows a type 1 extreme value distribution with scale parameter σ. A household chooses deductible d when $\mathcal{U}_i(d) > \mathcal{U}_i(d')$ for all $d' \neq d$, and thus the probability that a household chooses deductible d is

$$\Pr\left(\varepsilon_{i,d'} - \varepsilon_{i,d} < U(\mathcal{L}_i(d)) - U(\mathcal{L}_i(d'))\ \forall\ d' \neq d\right) = \frac{\exp\left(U(\mathcal{L}_i(d))/\sigma\right)}{\sum_{d'\in\mathcal{D}} \exp\left(U(\mathcal{L}_i(d'))/\sigma\right)}.$$

BMOT use these choice probabilities to construct the likelihood function and they estimate the model parameters by maximum likelihood. For the expected utility model, they report $\widehat{r} = 0.0129$. For the probability distortion model, they report $\widehat{r} = 0.00064$ and $\widehat{\Omega}(\mu_i) = \exp(-2.71 + 12.03\mu_i - 35.15\mu_i^2)$.

6.2.3 Results

We use the fitted models to investigate the extent to which home deductible choices would change if premiums were experience rated.[23] In particular, we examine for each model how the distribution of model-generated deductible choices in the home sample changes when we move from *prior premiums* $\widehat{p}_i$ to *univariate premiums* $\widehat{p}_i^{UV}$ and *multivariate premiums* $\widehat{p}_i^{MV}$. As before, $\widehat{p}_i$ denotes premiums calculated using prior claim probabilities, $\widehat{\mu}_i = 1 - \exp(-\widehat{\lambda}_i)$; $\widehat{p}_i^{UV}$ denotes premiums calculated using univariate posterior claim probabilities, $\widehat{\mu}_i^{UV} = 1 - \exp(-\widehat{\vartheta}_i)$, and reflects experience rating within coverages; and $\widehat{p}_i^{MV}$ denotes premiums calculated using multivariate posterior claim probabilities, $\widehat{\mu}_i^{MV} = 1 - \exp(-\widehat{\theta}_i)$, and reflects experience rating within and across coverages.[24]

Table 9 presents the results. In addition to displaying the distributions, it reports for each model the percentage of policies in which the deductible choice changes when we move from prior premiums to univariate and multivariate premiums, as well as the resulting changes in coverage. The main takeaway is that the response of deductible choices to experience rating is substantial under each model. Under the expected utility model, experience rating induces a large response: when we move from $\widehat{p}_i$ to $\widehat{p}_i^{UV}$ and $\widehat{p}_i^{MV}$, the deductible choice changes in 13 to 14 percent of policies, resulting in an aggregate change in coverage of \$2.5 to \$2.7 million, or roughly \$314 per policy (among policies with a change). Under the probability distortion model, experience rating induces a smaller, but still large, response: when we move from $\widehat{p}_i$ to $\widehat{p}_i^{UV}$ and $\widehat{p}_i^{MV}$, the deductible choice changes in 11 percent of policies, resulting in an aggregate change in coverage of \$2.1 to \$2.2 million, or roughly \$310 per policy (among policies with a change).

Two important insights follow from the results. First, a failure to experience rate premiums can lead to sizable distortions in households' demand

[23] Despite the simplistic assumption of homogeneous preferences, both models fit the data reasonably well: the expected utility model correctly predicts the deductible choice for 53 percent of the policies and the probability distortion model correctly predicts the deductible choice for 62 percent of the policies.

[24] In each case, we assume households believe their claim probability is $\mu_i = 1 - \exp(-\widehat{\theta}_i)$.

for insurance coverage. Second, the magnitude of such demand distortions depends on the nature of households' risk preferences, and in particular on the sources of their risk aversion. We expand upon these points in the next section, where we discuss legal restrictions on experience rating and what our findings imply about their economic consequences.

7 Economics of Legal Restrictions on Experience Rating

In the United States, states generally impose legal restrictions on the ability of insurers to engage in experience rating within and across lines of coverage. Under New York law, for example, although premiums for auto collision coverage may be experience rated, premiums for auto comprehensive and home all perils coverages may not be experience rated; moreover, neither auto comprehensive nor home claims may be used to experience rate auto collision premiums.[25] The principal justifications for legal restrictions on experience rating and other forms of risk classification are equity considerations (distributional and deontological) (e.g., Abraham 1985, 1986; Avraham et al. 2014), including, for instance, access to insurance for high risk, low income insureds (e.g., Meier 1991; Thiery and Van Schoubroeck 2006; Thomas 2007; Dionne and Rothchild 2014). A principal criticism of such legal restrictions is that they lead to unpriced heterogeneity and in turn to distortions in households' demand for insurance.

Our findings provide two key insights with respect to this criticism of legal restrictions on experience rating in insurance markets. First, our findings support the criticism—they suggest that a failure to experience rate premiums leads to material pricing distortions which, in turn, can lead to sizeable demand distortions. Moreover, our findings reveal that risk type is correlated across lines of coverage and document the incremental effects of restricting cross-coverage experience rating.

[25] See N.Y. Comp. Codes R. & Regis. tit. 11, §§ 161.1, 161.8 & 169.1 (2012); N.Y. Ins. Dep't OGC Op. No. 08-07-16 (2008).

Second, our findings suggest that the magnitude of the demand effects of legal restrictions on experience rating depends on the nature of insureds' risk preferences. In particular, we find that the demand distortions arising from a failure to experience rate premiums are greater if we model insureds according to expected utility theory than if we model them using a generalization of the expected utility model that allows for probability distortions. In both models, an insured's demand for insurance is driven by its aversion to risk. The key difference between the models is the source of the insured's aversion to risk. Under the expected utility model, the source is "standard" risk aversion, i.e., diminishing marginal utility for wealth, which is captured by a concave Bernoulli utility function. Under the probability distortion model, the source is not only standard risk aversion but also nonstandard sources such as probability weighting, loss aversion, or disappointment aversion, which are captured in our setting by a nonlinear probability distortion function (Barseghyan et al. 2013). Stated another way, in the expected utility model demand for insurance is driven solely by nonlinear evaluation of wealth, whereas in the probability distortion model demand for insurance is driven also by nonlinear evaluation of risk. Consequently, given any level of overall aversion to risk, the price elasticity of demand is greater (in absolute terms) under the expected utility model than it is under the probability distortion model. Intuitively speaking, this is because in the expected utility model overall aversion to risk is concentrated in the component of preferences that is sensitive to price changes (the Bernoulli utility function) whereas in the probability distortion model overall aversion to risk is distributed in part to the component of preferences that is insensitive to price changes (the probability distortion function).[26]

To illustrate how price elasticity depends on the source of aversion to risk, consider an insured whose risk preferences are represented by equation (2). This implies that the insured's price elasticity of utility is $\partial U/\partial p(d) = -1 - r[p(d) + \Omega(\mu)d]$. Now suppose that the insured has a claim probability $\mu =$

[26] Recent empirical work on risk preferences suggests that insureds' aversion to risk arises in part from sources other than diminishing marginal utility for wealth (Sydnor 2010; Barseghyan et al. 2011), and that probability distortions play an important and perhaps leading role (Barseghyan et al. 2013, 2014).

0.05 and faces a premium of \$200 for coverage with a deductible of \$1,000. Suppose further that the insured's maximum willingness to pay (WTP) to reduce its deductible from \$1,000 to \$500 is fifty dollars. Note that this WTP is a measure of the insured's overall aversion to risk, and that multiple combinations of r and $\Omega(0.05)$ are consistent with this WTP. For example, the following combinations are both consistent with this WTP: (i) $r = 0.00222$ and $\Omega(0.05) = 0.05$ (standard risk aversion only) and (ii) $r = 0$ and $\Omega(0.05) = 0.10$ (probability distortions only). However, the price elasticity of utility is 1.555 under combination (i) and only 1.0 under combination (ii). More generally, given this WTP, the insured's price elasticity of utility (and hence of demand) will be greater the greater is r (and thus the lesser is $\Omega(\mu)$). More generally still, given any level of overall aversion to risk, an insured's price elasticity of demand will be greater the greater is the extent to which its overall aversion to risk is attributable to standard risk aversion and the lesser is the extent to which its overall risk aversion is attributable to other sources that manifest as probability distortions.

8 Conclusion

We explore the information value of insureds' claims experience and the economic effects of experience rating insurance premiums. Using data on households' claims histories in auto and home insurance, we first estimate the variance-covariance matrix of unobserved heterogeneity in claim risk. Among other things, we find that unobserved heterogeneity is positively correlated across coverages. Next, we utilize the estimates to update a priori predictions about the households' claim risk. We show that conditioning on claims experience leads to material refinements of the households' predicted claim rates. We also demonstrate the incremental value of conditioning across lines of coverage (as well as within lines of coverage) in terms of updating the household's predicted claim rates and improving their accuracy. We then show that experience rating within and across lines of coverage leads to material refinements in the households' home premiums. Finally, we assess the extent to which the households' demand for home coverage would respond to experience rating under two models of risky choice: the standard expected utility model and a generalization of the standard model that features probability distortions. We find that the demand response to experience rating is large under both models but smaller under the probability distortion model. We conclude that legal restrictions on experience rating have the potential to lead to unpriced heterogeneity and demand distortions, but that the magnitude of the demand distortions depends on the nature of insureds' risk preferences.

In addition to the foregoing contributions, the paper has a noteworthy connection to recent work on the stability of risk preferences across contexts. In particular, our finding that unobserved heterogeneity in claim risk is positively correlated across coverages suggests that there is a domain-general component to risk type. This complements existing and ongoing research which suggests that risk preferences, though not completely stable across contexts (Barseghyan et al. 2011), also have a domain-general component (e.g., Barksy et al. 1997; Dohmen et al. 2011; Einav et al. 2012; Barseghyan et al. 2014).

References

Abraham, K. S. (1985): "Efficiency and Fairness in Insurance Risk Classification," *Virginia Law Review*, 71, 403–451.

——— (1986): *Distributing Risk: Insurance, Legal Theory, and Public Policy*, New Haven, CT: Yale University Press.

Angers, J.-F., D. Desjardins, G. Dionne, and F. Guertin (2006): "Vehicle and Fleet Random Effects in a Model of Insurance Rating for Fleets of Vehicles," *ASTIN Bulletin*, 36, 25–77.

Antonio, K., M. Guillén, and A.-M. P. M. Martin (2011): "Multidimensional Credibility: A Bayesian Analysis of Policholders Holding Multiple Contracts," Working Paper, University of Amsterdam.

Antonio, K. and E. A. Valdez (2012): "Statistical Concepts of *A Priori* and *A Posteriori* Risk Classification in Insurance," *AstA Advances in Statistical Analysis*, 96, 187–224.

Avraham, R., K. D. Logue, and D. Schwarcz (2014): "Understanding Insurance Anti-Discrimination Laws," *Southern California Law Review*, 87, 195–274.

Barksy, R. B., F. T. Juster, M. S. Kimball, and M. D. Shapiro (1997): "Preference Parameters and Behavioral Heterogeneity: An Experimental Approach in the Health and Retirement Study," *Quarterly Journal of Economics*, 112, 537–579.

Barseghyan, L., F. Molinari, T. O'Donoghue, and J. C. Teitelbaum (2013): "The Nature of Risk Preferences: Evidence from Insurance Choices," *American Economic Review*, 103, 2499–2529.

Barseghyan, L., F. Molinari, and J. C. Teitelbaum (2014): "Inference under Stability of Risk Preferences," Working Paper 1344, Georgetown University Law Center.

Barseghyan, L., J. Prince, and J. C. Teitelbaum (2011): "Are Risk Preferences Stable Across Contexts? Evidence from Insurance Data," *American Economic Review*, 101, 591–631.

Buchmueller, T. and J. DiNardo (2002): "Did Community Rating Induce an Adverse Selection Death Spiral? Evidence from New York, Pennsylvania, and Connecticut," *American Economic Review*, 92, 280–294.

Bundorf, M. K., J. Levin, and N. Mahoney (2012): "Pricing and Welfare in Health Plan Choice," *American Economic Review*, 102, 3214–3248.

Bundorf, M. K. and K. I. Simon (2006): "The Effects of Rate Regulation on Demand for Supplemental Health Insurance," *American Economic Review: Papers and Proceedings*, 96, 67–71.

Bühlmann, H. and A. Gisler (2005): *A Course in Credibility Theory and its Applications*, Berlin: Springer.

Cohen, A. and P. Siegelman (2010): "Testing for Adverse Selection in Insurance Markets," *Journal of Risk and Insurance*, 77, 39–84.

Crocker, K. J. and A. Snow (1986): "The Efficiency Effects of Categorical Discrimination in the Insurance Industry," *Journal of Political Economy*, 94, 321–344.

——— (2011): "Multidimensional Screening in Insurance Markets with Adverse Selection," *Journal of Risk and Insurance*, 78, 287–307.

Denuit, M., X. Maréchal, S. Pitrebois, and J.-F. Walhin (2007): *Actuarial Modelling of Claim Counts: Risk Classification, Credibility and Bonus-Malus Systems*, Chichester: John Wiley.

Desjardins, D., G. Dionne, and J. Pinquet (2001): "Experience Rating Schemes for Fleets of Vehicles," *ASTIN Bulletin*, 31, 81–105.

Dionne, G. and C. Rothchild (2014): "Economic Effects of Risk Classification Bans," *Geneva Risk and Insurance Review*, 39, 184–221.

Dohmen, T., A. Falk, D. Huffman, U. Sunde, J. Schupp, and G. G. Wagner (2011): "Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences," *Journal of the European Economic Association*, 9, 522–550.

Eeckhoudt, L., C. Gollier, and H. Schlesinger (2005): *Economic and Financial Decisions under Risk*, Princeton, NJ: Princeton University Press.

Einav, L., A. Finkelstein, and M. R. Cullen (2010): "Estimating Welfare in Insurance Markets Using Variation in Prices," *Quarterly Journal of Economics*, 125, 877–921.

Einav, L., A. Finkelstein, I. Pascu, and M. R. Cullen (2012): "How General are Risk Preferences? Choice under Uncertainty in Different Domains," *American Economic Review*, 102, 2606–2638.

Englund, M., M. Guillén, J. Gustafsson, J. P. Nielsen, and F. Thuring (2009): "Multidimensional Credibility with Time Effects: An Application to Commerical Business Lines," *Journal of Risk and Insurance*, 76, 443–453.

Englund, M., M. Guillén, J. Gustafsson, L. H. Nielsen, and J. P. Nielsen (2008): "Multivariate Latent Risk: A Credibility Approach," *ASTIN Bulletin*, 38, 137–146.

Finkelstein, A., J. Poterba, and C. Rothschild (2009): "Redistribution by Insurance Market Regulation: Analyzing a Ban on Gender-Based Retirement Annuities," *Journal of Financial Economics*, 91, 38–58.

Frees, E. J., G. Meyers, and A. D. Cummings (2010): "Dependent Multi-Peril Ratemaking Models," *ASTIN Bulletin*, 40, 699–726.

Geruso, M. (2013): "Selection in Employer Health Plans: Homogeneous Prices and Heterogeneous Preferences," Working Paper, University of Texas at Austin, Department of Economics.

Gouriéroux, C., A. Monfort, and A. Trognon (1984a): "Psuedo Maximum Likelihood Methods," *Econometrica*, 52, 681–700.

——— (1984b): "Psuedo Maximum Likelihood Methods: Applications to Poisson Models," *Econometrica*, 52, 701–720.

Gul, F. (1991): "A Theory of Disappointment Aversion," *Econometrica*, 59, 667–686.

Hoy, M. (1982): "Categorizing Risks in the Insurance Industry," *Quarterly Journal of Economics*, 97, 321–336.

——— (2006): "Risk Classification and Social Welfare," *Geneva Papers on Risk and Insurance: Issues and Practice*, 31, 245–269.

Jewell, W. S. (1974): "Exact Multidimensional Credibility," *Bulletin of Swiss Association of Actuaries*, 74, 193–214.

Kőszegi, B. and M. Rabin (2006): "A Model of Reference-Dependent Preferences," *Quarterly Journal of Economics*, 121, 1133–1166.

——— (2007): "Reference-Dependent Risk Attitudes," *American Economic Review*, 97, 1047–1073.

Lemaire, J. (1995): *Bonus-Malus Systems in Automobile Insurance*, Boston: Kluwer.

Liang, K.-Y. and S. L. Zeger (1986): "Longitudinal Data Analysis Using Generalized Linear Models," *Biometrika*, 73, 13–22.

Meier, K. J. (1991): "The Politics of Insurance Regulation," *Journal of Risk and Insurance*, 58, 700–713.

Morris, D. S. (2012): "A Semiparametric Approach for Multivariate Longitudinal Count Data," in *Methods for Multivariate Longitudinal Count and Duration Models with Applications in Economics*, Ph.D. Dissertation, Cornell University, chap. 2.

Pinquet, J. (1998): "Designing Optimal Bonus-Malus Systems from Different Types of Claims," *ASTIN Bulletin*, 28, 205–220.

——— (2000): "Experience Rating Through Heterogeneous Models," in *Handbook of Insurance*, ed. by G. Dionne, Boston: Kluwer, chap. 14.

——— (2013): "Experience Rating in Non-Life Insurance," in *Handbook of Insurance*, Second Edition, ed. by G. Dionne, New York: Springer, chap. 17.

Prentice, R. L. (1988): "Correlated Binary Regression with Covariates Specific to Each Binary Observation," *Biometrics*, 44, 1033–1048.

Quiggin, J. (1982): "A Theory of Anticipated Utility," *Journal of Economic Behavior and Organization*, 3, 323–343.

Rothschild, C. (2011): "The Efficiency of Categorical Discrimination in Insurance Markets," *Journal of Risk and Insurance*, 78, 267–285.

Salanié, B. (1997): *The Economics of Contracts: A Primer*, Cambridge, MA: MIT Press.

Schlesinger, H. (2000): "The Theory of Insurance Demand," in *Handbook of Insurance*, ed. by G. Dionne, Boston: Kluwer, chap. 5.

Simon, K. I. (2005): "Adverse Selection in Health Insurance Markets? Evidence from State Small-Group Health Insurance Reforms," *Journal of Public Economics*, 89, 1865–1877.

Sydnor, J. (2010): "(Over)Insuring Modest Risks," *American Economic Journal: Applied Economics*, 2, 177–199.

Thiery, Y. and C. Van Schoubroeck (2006): "Fairness and Equality and Insurance Classification," *Geneva Papers on Risk and Insurance: Issues and Practice*, 31, 190–211.

Thomas, R. G. (2007): "Some Novel Perspectives on Risk Classification," *Geneva Papers on Risk and Insurance: Issues and Practice*, 32, 105–132.

——— (2008): "Loss Coverage as a Public Policy Objective for Risk Classification Schemes," *Journal of Risk and Insurance*, 75, 997–1018.

Vuong, Q. H. (1989): "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses," *Econometrica*, 57, 307–333.

Wooldridge, J. M. (2002): *Econometric Analysis of Cross Section and Panel Data*, Cambridge, MA: MIT Press.

Zeger, S. L. and K.-Y. Liang (1986): "Longitudinal Data Analysis for Discrete and Continuous Outcomes," *Biometrics*, 42, 121–130.

Zeger, S. L., K.-Y. Liang, and P. S. Albert (1988): "Models for Longitudinal Data: A Generalized Estimating Equation Approach," *Biometrics*, 44, 1049–1060.

Ziegler, A. (2011): *Generalized Estimating Equations*, New York: Springer.

Table 1: Summary of Claims, Premiums, and Deductibles
Tricoverage Sample (294,917 household-years)

	Mean	Standard deviation	Minimum	Maximum
Claims (count):				
Auto collision	0.107	0.334	0	5
Auto comprehensive	0.032	0.188	0	5
Home	0.079	0.299	0	6
Premiums (dollars):				
Auto collision	200	104	20	2,520
Auto comprehensive	127	70	6	2,524
Home	548	309	50	10,224
Deductibles (dollars):				
Auto collision	396	181	100	1,000
Auto comprehensive	273	176	50	1,000
Home	350	242	100	5,000

Table 2: Association Parameter Estimates
Tricoverage Sample (294,917 household-years)

	Estimate	95 percent confidence interval	
Variances:			
Auto collision	0.107	0.065	0.149
Auto comprehensive	0.399	0.221	0.577
Home	0.405	0.383	0.428
Covariances:			
Auto collision and auto comprehensive	0.137	0.101	0.173
Auto collision and home	0.061	0.022	0.099
Auto comprehensive and home	0.225	0.179	0.271
Correlations:			
Auto collision and auto comprehensive	0.663	0.399	0.926
Auto collision and home	0.293	0.099	0.486
Auto comprehensive and home	0.559	0.389	0.729

Table 3: Association Parameter Estimates - Low and High Deductible Households

	Tricoverage sample (62,425 households; 294,917 household-years)			All deductibles ≤ $250 (22,072 households; 120,213 household-years)			Any deductible > $250 (40,353 households; 174,704 household-years)		
	Estimate	95 percent confidence interval		Estimate	95 percent confidence interval		Estimate	95 percent confidence interval	
Variances:									
Auto collision	0.107	0.065	0.149	0.094	0.038	0.150	0.108	0.051	0.166
Auto comprehensive	0.399	0.221	0.577	0.337	0.086	0.587	0.450	0.201	0.698
Home	0.405	0.383	0.428	0.388	0.281	0.496	0.246	0.038	0.454
Covariances:									
Auto collision and auto comprehensive	0.137	0.101	0.173	0.138	0.085	0.192	0.129	0.081	0.178
Auto collision and home	0.061	0.022	0.099	0.088	0.055	0.120	0.058	0.021	0.094
Auto comprehensive and home	0.225	0.179	0.271	0.224	0.157	0.290	0.217	0.152	0.282
Correlations:									
Auto collision and auto comprehensive	0.663	0.399	0.926	0.776	0.300	1.252	0.586	0.272	0.900
Auto collision and home	0.293	0.099	0.486	0.458	0.230	0.686	0.352	0.066	0.639
Auto comprehensive and home	0.559	0.389	0.729	0.619	0.312	0.926	0.652	0.270	1.034

Table 4: Descriptive Statistics for $\eta=(\theta-\lambda)/\lambda$
Tricoverage Sample (294,917 household-years)

Coverage	Prior claim rates	η<0 Observations	η<0 Mean	η<0 10th percentile	η<0 25th percentile	η>0 Observations	η>0 Mean	η>0 75th percentile	η>0 90th percentile
Auto collision	All	180,909	-0.066	-0.121	-0.093	114,008	0.101	0.140	0.227
	Low	46,914	-0.055	-0.100	-0.080	26,815	0.094	0.129	0.213
	Medium	89,988	-0.066	-0.120	-0.094	57,471	0.100	0.139	0.223
	High	44,007	-0.078	-0.143	-0.110	29,722	0.110	0.152	0.249
Auto Comprehensive	All	188,792	-0.132	-0.236	-0.186	106,125	0.231	0.310	0.531
	Low	47,384	-0.111	-0.198	-0.157	26,345	0.198	0.270	0.455
	Medium	94,742	-0.131	-0.233	-0.186	52,717	0.231	0.307	0.533
	High	46,666	-0.155	-0.273	-0.220	27,063	0.264	0.353	0.593
Home	All	196,205	-0.142	-0.241	-0.198	98,712	0.280	0.367	0.646
	Low	51,136	-0.120	-0.207	-0.166	22,593	0.273	0.350	0.630
	Medium	97,208	-0.146	-0.238	-0.200	50,251	0.277	0.365	0.642
	High	47,861	-0.159	-0.271	-0.227	25,868	0.292	0.390	0.663

Notes: A prior claim rate is "low" if it is in the bottom quartile and "high" if it is in the top quartile. It is "medium" otherwise. In the tricoverage sample, the respective low and high cutoffs are 0.078 and 0.127 in auto collision, 0.016 and 0.044 in auto comprehensive, and 0.054 and 0.096 in home.

Table 5: Descriptive Statistics for $\zeta=(\theta-\vartheta)/\vartheta$
Tricoverage Sample (294,917 household-years)

Coverage	Prior claim rates	ζ<0 Observations	ζ<0 Mean	ζ<0 10th percentile	ζ<0 25th percentile	ζ>0 Observations	ζ>0 Mean	ζ>0 75th percentile	ζ>0 90th percentile
Auto collision	All	198,557	-0.034	-0.059	-0.046	96,360	0.069	0.093	0.153
	Low	51,883	-0.030	-0.052	-0.041	21,846	0.067	0.092	0.148
	Medium	99,078	-0.034	-0.059	-0.047	48,381	0.069	0.093	0.153
	High	47,596	-0.038	-0.065	-0.052	26,133	0.070	0.095	0.157
Auto Comprehensive	All	183,662	-0.098	-0.173	-0.139	111,255	0.157	0.217	0.357
	Low	46,521	-0.095	-0.171	-0.135	27,208	0.163	0.224	0.370
	Medium	91,981	-0.099	-0.175	-0.140	55,478	0.159	0.220	0.359
	High	45,160	-0.098	-0.173	-0.139	28,569	0.149	0.207	0.334
Home	All	202,137	-0.044	-0.079	-0.059	92,780	0.093	0.132	0.211
	Low	51,610	-0.040	-0.074	-0.055	22,119	0.093	0.135	0.213
	Medium	101,001	-0.046	-0.082	-0.062	46,458	0.095	0.134	0.214
	High	49,526	-0.044	-0.080	-0.059	24,203	0.090	0.126	0.203

Notes: A prior claim rate is "low" if it is in the bottom quartile and "high" if it is in the top quartile. It is "medium" otherwise. In the tricoverage sample, the respective low and high cutoffs are 0.078 and 0.127 in auto collision, 0.016 and 0.044 in auto comprehensive, and 0.054 and 0.096 in home.

Table 6: Accuracy

Σ spec.	Baseline claim rates	Time horizon	Square root of MSE_k: Univariate Home	Univariate Coll	Univariate Comp	Multivariate Home	Multivariate Coll	Multivariate Comp	MSE_k^{MV}/MSE_k^{UV} Home	Coll	Comp	Percent of households for whom $MSE_{ik}^{MV} < MSE_{ik}^{UV}$ Home	Coll	Comp	Percent of iterations in which $MSE_{jk}^{MV} < MSE_{jk}^{UV}$ Home	Coll	Comp
A	Average	1	0.0439	0.0325	0.0188	0.0438	0.0324	0.0187	0.998	0.996	0.993	71.65	79.53	85.52	99.80	100.00	100.00
		5	0.0416	0.0319	0.0184	0.0412	0.0313	0.0179	0.990	0.984	0.971	88.60	96.72	98.67	100.00	100.00	100.00
		10	0.0392	0.0311	0.0179	0.0385	0.0302	0.0170	0.984	0.971	0.948	94.54	99.30	99.88	100.00	100.00	100.00
	Low	1	0.0340	0.0254	0.0101	0.0340	0.0253	0.0100	0.999	0.998	0.995	66.79	74.18	82.57	97.80	100.00	100.00
		5	0.0326	0.0250	0.0099	0.0324	0.0247	0.0097	0.994	0.989	0.976	83.71	93.06	97.70	100.00	100.00	100.00
		10	0.0310	0.0245	0.0098	0.0307	0.0240	0.0094	0.989	0.982	0.957	90.01	97.42	99.63	100.00	100.00	100.00
	High	1	0.0598	0.0412	0.0275	0.0596	0.0410	0.0273	0.997	0.995	0.991	76.01	84.48	89.06	99.80	100.00	100.00
		5	0.0557	0.0402	0.0266	0.0550	0.0393	0.0256	0.987	0.978	0.963	91.90	98.44	99.36	100.00	100.00	100.00
		10	0.0515	0.0389	0.0256	0.0505	0.0375	0.0240	0.979	0.963	0.938	96.18	99.71	99.90	100.00	100.00	100.00
B	Average	1	0.0347	0.0494	0.0149	0.0346	0.0493	0.0148	0.998	0.998	0.992	73.87	73.38	89.76	100.00	99.80	100.00
		5	0.0335	0.0471	0.0147	0.0332	0.0466	0.0142	0.991	0.990	0.964	90.86	91.25	99.63	100.00	100.00	100.00
		10	0.0323	0.0447	0.0145	0.0317	0.0439	0.0136	0.984	0.983	0.936	95.81	96.27	99.98	100.00	100.00	100.00
	Low	1	0.0268	0.0386	0.0080	0.0268	0.0385	0.0079	0.999	0.999	0.994	69.88	68.85	86.63	99.40	99.40	100.00
		5	0.0261	0.0372	0.0079	0.0260	0.0370	0.0077	0.994	0.994	0.971	86.81	86.27	98.97	100.00	100.00	100.00
		10	0.0253	0.0356	0.0078	0.0250	0.0352	0.0074	0.989	0.989	0.947	92.40	92.06	99.93	100.00	100.00	100.00
	High	1	0.0473	0.0625	0.0219	0.0472	0.0623	0.0216	0.997	0.997	0.990	76.86	76.87	91.89	100.00	100.00	100.00
		5	0.0453	0.0590	0.0214	0.0447	0.0581	0.0204	0.988	0.986	0.955	93.07	94.10	99.87	100.00	100.00	100.00
		10	0.0430	0.0551	0.0209	0.0422	0.0538	0.0193	0.980	0.977	0.924	97.19	98.03	99.99	100.00	100.00	100.00

Notes: In specification A, we set Σ to match the estimates from the data. In specification B, we fix the within-coverage variances such that each equals 0.250 and we adjust the cross-coverage covariances such that the cross-coverage correlations still match the estimates from the data. Average baseline claim rates roughly correspond to the mean prior claim rates in the tricoverage sample: λ=(0.100,0.030,0.070). Low baseline claim rates correspond to the 25th percentiles: λ=(0.078,0.016,0.054). High baseline claim rates correspond to the 75th percentiles: λ=(0.127,0.044,0.096).

Table 7: Fitted Premiums and Implied Risk Elasticities of Price
Home Sample (62,425 policies)

	Fitted value of $\bar{p}_i$	Implied value of ξ_i
Mean	356	0.82
Standard deviation	309	0.57
5th percentile	108	0.51
10th percentile	152	0.63
25th percentile	223	0.77
Median	309	0.84
75th percentile	423	0.94
90th percentile	576	0.97
95th percentile	712	1.00

Table 8: Descriptive Statistics for $\Gamma=(p^{MV}-p)/p$ and $\Delta=(p^{MV}-p^{UV})/p^{UV}$
Home Sample (62,425 policies)

Prior claim rates	Policies	Mean	10th percentile	25th percentile	Policies	Mean	75th percentile	90th percentile
	Γ<0				Γ>0			
All	43,814	-0.110	-0.208	-0.155	18,611	0.252	0.328	0.582
Low	12,106	-0.071	-0.129	-0.104	3,496	0.252	0.320	0.589
Medium	21,353	-0.120	-0.206	-0.167	9,865	0.251	0.326	0.572
High	10,355	-0.134	-0.248	-0.202	5,250	0.254	0.342	0.591
	Δ<0				Δ>0			
All	44,669	-0.034	-0.066	-0.047	17,756	0.085	0.126	0.194
Low	11,930	-0.025	-0.049	-0.035	3,672	0.080	0.127	0.189
Medium	22,078	-0.038	-0.072	-0.052	9,140	0.090	0.129	0.203
High	10,661	-0.036	-0.070	-0.050	4,944	0.079	0.115	0.184

Notes: A prior claim rate is "low" if it is in the bottom quartile and "high" if it is in the top quartile. It is "medium" otherwise. In the home sample, the low and high cutoffs are .063 and .109, respectively.

Table 9: Response of Deductible Choices to Experience Rating
Home Sample (62,425 policies)

	Expected Utility Model			Probability Distortion Model		
Deductible Choice	Prior Premiums	Univariate Premiums	Multivariate Premiums	Prior Premiums	Univariate Premiums	Multivariate Premiums
$100	1,610	1,590	1,589	1,914	1,965	1,988
$250	37,319	38,555	38,823	46,210	46,595	46,629
$500	17,640	16,744	16,544	11,233	10,699	10,599
$1,000	5,745	5,439	5,372	3,007	3,099	3,137
$2,500	80	74	75	30	37	43
$5,000	31	23	22	31	30	29
Policies with change in deductible (percent)	-	12.85	13.85	-	10.64	11.32
Aggregate change in coverage (dollars)	-	2,520,400	2,710,100	-	2,058,150	2,192,150
Per policy change in coverage (dollars), policies with change	-	314.26	313.56	-	309.87	310.28
Per policy change in coverage (dollars), all policies	-	40.37	43.41	-	32.97	35.12

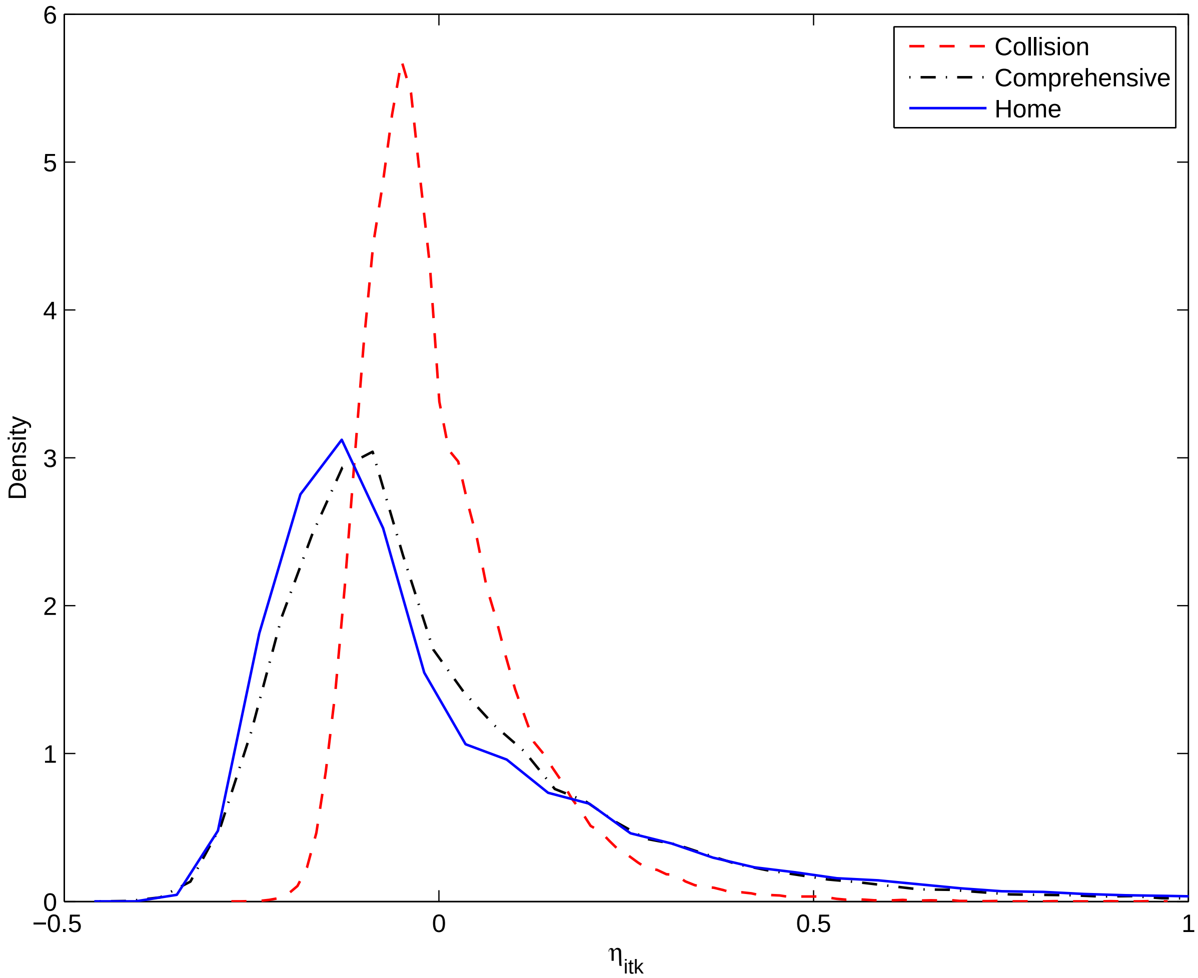

Figure 1: Kernel Density of η_{itk}

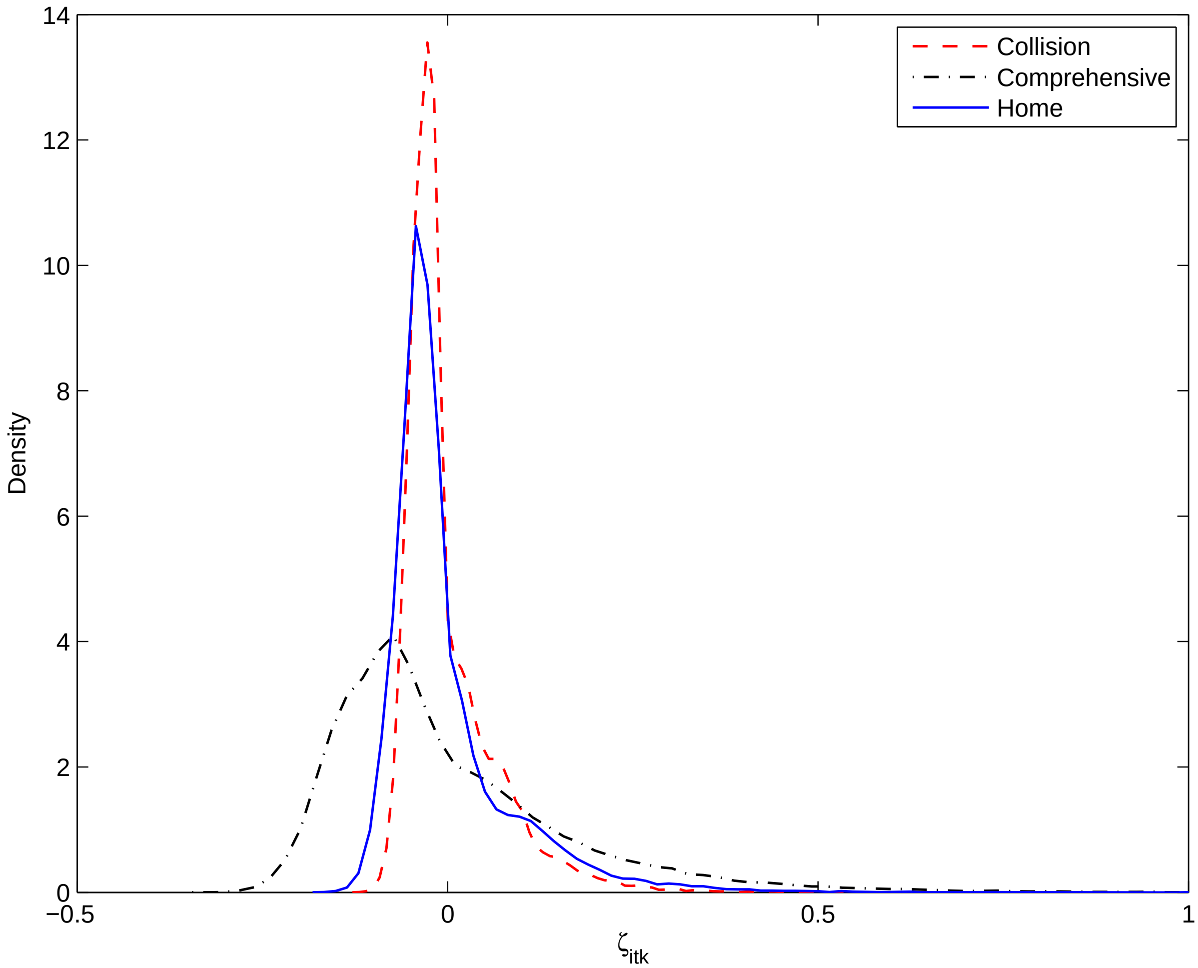

Figure 2: Kernel Density of ζ_{itk}

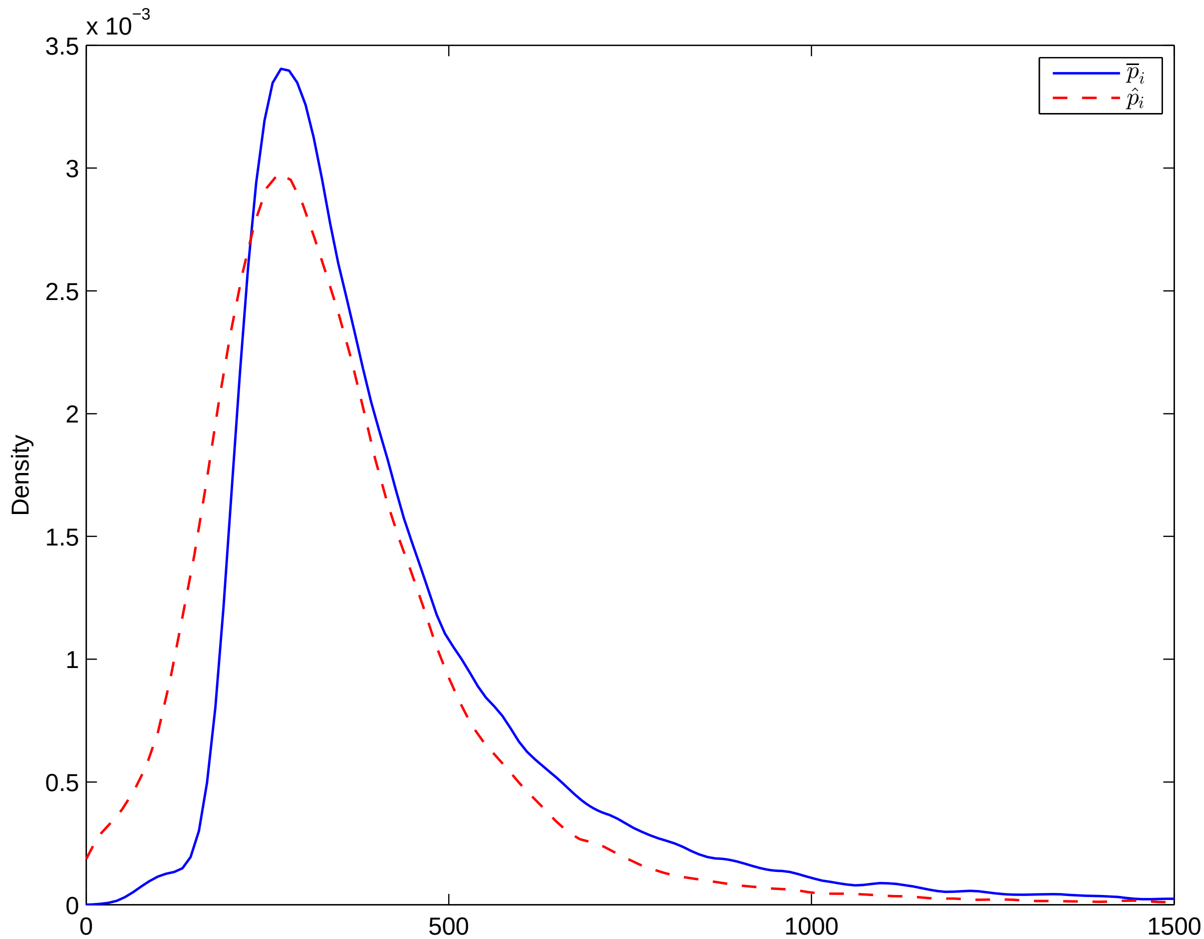

Figure 3: Kernel Densities of $\overline{p}_i$ and $\widehat{p}_i$

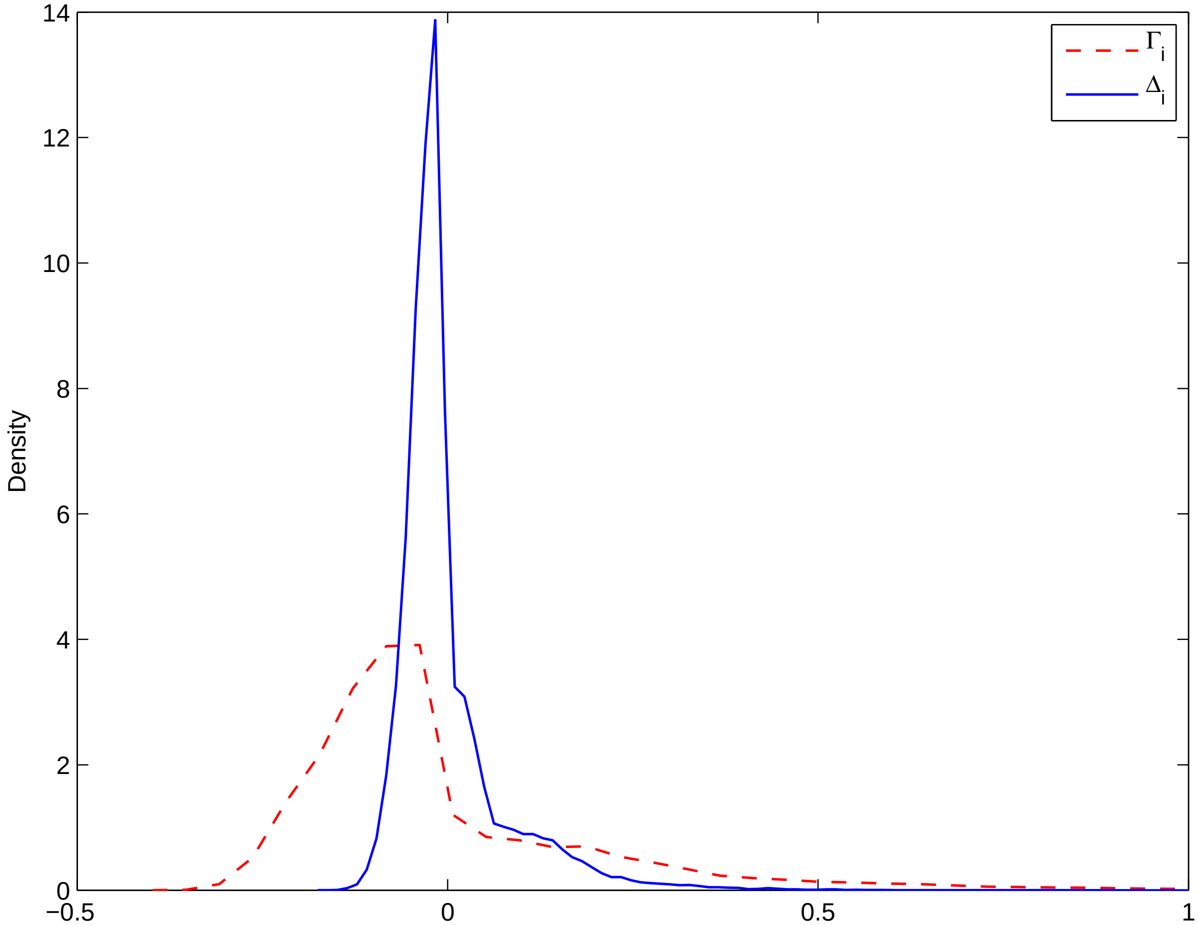

Figure 4: Kernel Densities of Γ_i and Δ_i

Appendix
to
Unlucky or Risky? Unobserved Heterogeneity and Experience Rating in Insurance Markets

Levon Barseghyan
Cornell University

Francesca Molinari
Cornell University

Darcy Steeg Morris
U.S. Census Bureau

Joshua C. Teitelbaum
Georgetown University

Draft: September 30, 2014

A Regression Parameter Estimates

Tables A5 and A6 report the estimates of the regression parameters, $\boldsymbol{\beta}$. Although $\boldsymbol{\beta}$ is not the object of principal interest, the estimates reveal several noteworthy facts. First, auto claim rates (collision and comprehensive) are negatively related to insurance score (which is based on information contained in credit reports) but positively related to the age and number of vehicles. However, they are not correlated with vehicle safety features (passive restraint, anti-theft, and anti-lock brakes). Second, collision claim rates are negatively related to the age of the primary driver and are higher for households in which the primary driver is female. Conversely, comprehensive claim rates are positively related to the age of the primary driver and are lower for households in which the primary driver is female. Third, collision claim rates are higher for households with three or more drivers. Finally, home claim rates are negatively related to insurance score but positively related to the age and insured value of the home. In addition, they are higher for homes that are used for farming or business and for homes that are not the owner's primary residence. Home claim rates, however, are not correlated with home safety features (masonry construction, distance to fire hydrant, and alarm or other protection).

B Estimation Strategy: Marginal Moments and Estimating Equations

Let y_{itk} denote the number of claims for household i in year t under coverage k, where $i = 1, \ldots, N$, $t = 1, \ldots, T_i$, and $k \in \{c, m, h\}$. Similarly, let $\mathbf{x}_{itk}$ denote a vector of observables (plus a constant) for household i in year t under coverage k. Let λ_{itk} denote household i's baseline claim rate in year t under coverage k, and let ϵ_{ik} denote a time-constant random effect for household i under coverage k. Also, let $\mathbf{y}_{ik} \equiv (y_{i1k}, ..., y_{iT_ik})$ and $\mathbf{y}_i \equiv (\mathbf{y}_{ic}, \mathbf{y}_{im}, \mathbf{y}_{ih})$, and let $\boldsymbol{\lambda}_{ik} \equiv (\lambda_{i1k}, ..., \lambda_{iT_ik})$ and $\boldsymbol{\lambda}_i \equiv (\boldsymbol{\lambda}_{ic}, \boldsymbol{\lambda}_{im}, \boldsymbol{\lambda}_{ih})$.

The first two marginal moments for the class of models used in this reserach—longitudinal multivariate count models with multiplicative correlated random effects—are

$$E(y_{itk}|\mathbf{x}_{itk}) = \exp\left(\mathbf{x}'_{itk}\boldsymbol{\beta}_k\right) = \lambda_{itk}$$

and

$$\mathbf{V}_i \equiv Var(\mathbf{y}_i|\mathbf{x}_i) = diag(\boldsymbol{\lambda}'_i) + \boldsymbol{\Sigma} \otimes \mathbf{1}_{T_i}\mathbf{1}'_{T_i} \circ \boldsymbol{\lambda}_i\boldsymbol{\lambda}'_i,$$

where $\circ$ is element-wise multiplication, $\otimes$ is the Kronecker product, and $\mathbf{1}_{T_i}$ is a T_i-dimensional vector of ones.

The moment-based approach for fitting this model relies on the moment conditions implied by the marginal mean and variance along with the basic assumptions for multiplicative correlated random effects models. The estimator $(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Sigma}})$ for $\boldsymbol{\beta} \equiv [\boldsymbol{\beta}_c \; \boldsymbol{\beta}_m \; \boldsymbol{\beta}_h]'$ and $\boldsymbol{\Sigma}$ is defined as the solution to

$$\sum_i \begin{pmatrix} \mathbf{D}'_i & 0 \\ 0 & \mathbf{E}'_i \end{pmatrix} \begin{pmatrix} \mathbf{V}_i & 0 \\ 0 & \mathbf{I} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{y}_i - \boldsymbol{\lambda}_i \\ \mathbf{R}^*_i - \mathbf{V}^*_i \end{pmatrix} = 0,$$

where $\mathbf{D}_i \equiv \frac{\partial \boldsymbol{\lambda}_i}{\partial \boldsymbol{\beta}} = diag\left[\mathbf{x}'_{ic}\boldsymbol{\lambda}_{ic} \; \mathbf{x}'_{im}\boldsymbol{\lambda}_{im} \; \mathbf{x}'_{ih}\boldsymbol{\lambda}_{ih}\right]'$, $\mathbf{V}_i$ is the model based variance-covariance matrix as defined above,

$$\begin{aligned} \mathbf{E}_i &\equiv \frac{\partial \mathbf{V}^*_i}{\partial \boldsymbol{\Sigma}^*} \\ &= diag\left[(\boldsymbol{\lambda}_{ic}\boldsymbol{\lambda}'_{ic})^* \;\; (\boldsymbol{\lambda}_{im}\boldsymbol{\lambda}'_{im})^* \;\; (\boldsymbol{\lambda}_{ih}\boldsymbol{\lambda}'_{ih})^* \;\; (\boldsymbol{\lambda}_{ic}\boldsymbol{\lambda}'_{im})^* \;\; (\boldsymbol{\lambda}_{ic}\boldsymbol{\lambda}'_{ih})^* \;\; (\boldsymbol{\lambda}_{im}\boldsymbol{\lambda}'_{ih})^*\right], \end{aligned}$$

$\mathbf{I}$ is the identity matrix, and $\mathbf{R}_i$ is the cross product of residuals $r_{itk} \equiv y_{itk} - \lambda_{itk}$. Also, let * indicate a half-vectorization operator, such that $\mathbf{R}^*_i$, $\mathbf{V}^*_i$, and $\boldsymbol{\Sigma}^*$ are the vectors of the upper triangular elements of the matrices $\mathbf{R}_i$, $\mathbf{V}_i$, and $\boldsymbol{\Sigma}$, respectively. The roots of the set of estimating equations are obtained via an iterative procedure, updated at each iteration with the consistent estimator of $\boldsymbol{\beta}$ given $\widehat{\boldsymbol{\Sigma}}$ and the consistent estimator of $\boldsymbol{\Sigma}$ given $\widehat{\boldsymbol{\beta}}$, until convergence. See Morris (2012) for more details on the estimation algorithm and asymptotic results for joint inference.

C Derivations of $E^{UV}(\epsilon_{ik}|\mathbf{y}_{ik})$ and $E^{MV}(\epsilon_{ik}|\mathbf{y}_i)$

Let y_{itk} denote the number of claims for household i in year t under coverage k, where $i = 1, \ldots, N$, $t = 1, \ldots, T_i$, and $k \in \{c, m, h\}$. Similarly, let $\mathbf{x}_{itk}$ denote a vector of observables (plus a constant) for household i in year t under coverage k. Let λ_{itk} denote household i's baseline claim rate in year t under coverage k, and let ϵ_{ik} denote a time-constant random effect for household i under coverage k. Also, let $\mathbf{y}_{ik} \equiv (y_{i1k}, ..., y_{iT_ik})$ and $\mathbf{y}_i \equiv (\mathbf{y}_{ic}, \mathbf{y}_{im}, \mathbf{y}_{ih})$.

C.1 Derivation of $E^{MV}(\epsilon_{ik}|\mathbf{y}_i)$

We assume

$$y_{itk}|\mathbf{x}_{itk} \sim Poisson\left(\lambda_{itk}\epsilon_{ik}\right),$$

where

$$\lambda_{itk} = \exp\left(\mathbf{x}'_{itk}\boldsymbol{\beta}_k\right)$$

and $\boldsymbol{\epsilon}_i \equiv [\epsilon_{ic}\ \epsilon_{im}\ \epsilon_{ih}]' \overset{iid}{\sim} lognormal$ with $E(\boldsymbol{\epsilon}_i) = [1\ 1\ 1]'$ and $V(\boldsymbol{\epsilon}_i) = \boldsymbol{\Sigma}$. This leads to the following probability distribution functions:

$$\begin{aligned} f(\mathbf{y}_i|\boldsymbol{\epsilon}_i) &= \prod_k \prod_t Poisson(\lambda_{itk}\epsilon_{ik}) \\ &= \prod_k \prod_t \frac{(\epsilon_{ik}\lambda_{itk})^{y_{itk}}}{y_{itk}!} e^{-\epsilon_{ik}\lambda_{itk}} \\ &= \left(\prod_k \epsilon_{ik}^{\sum_t y_{itk}} e^{-\epsilon_{ik}\sum_t \lambda_{itk}}\right)\left(\prod_k \prod_t \frac{\lambda_{itk}^{y_{itk}}}{y_{itk}!}\right), \end{aligned}$$

$$\begin{aligned} f(\tilde{\epsilon}_i) &= Normal(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Sigma}}) \\ &= \frac{1}{(2\pi)^{3/2}}|\tilde{\boldsymbol{\Sigma}}|^{-1/2} e^{-\frac{1}{2}(\tilde{\epsilon}_i - \tilde{\boldsymbol{\mu}})'\tilde{\boldsymbol{\Sigma}}^{-1}(\tilde{\epsilon}_i - \tilde{\boldsymbol{\mu}})}, \end{aligned}$$

where $\tilde{\epsilon}_i \equiv \ln(\epsilon_i)$, $\tilde{\boldsymbol{\mu}} \equiv -\frac{diag(\tilde{\boldsymbol{\Sigma}})}{2}$, and $\tilde{\boldsymbol{\Sigma}} \equiv \ln(\boldsymbol{\Sigma}+1)$, and

$$\begin{aligned} f(\mathbf{y}_i) &= \int_{\tilde{\epsilon}_{ic}} \int_{\tilde{\epsilon}_{im}} \int_{\tilde{\epsilon}_{ih}} f(\mathbf{y}_i|\tilde{\epsilon}_i) f(\tilde{\epsilon}_i)\; d\tilde{\epsilon}_{ih} d\tilde{\epsilon}_{im} d\tilde{\epsilon}_{ic} \\ &= \int_{\tilde{\epsilon}_{ic}} \int_{\tilde{\epsilon}_{im}} \int_{\tilde{\epsilon}_{ih}} \prod_k \prod_t Poisson(\lambda_{itk} e^{\tilde{\epsilon}_{ik}})\; MVN(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Sigma}})\; d\tilde{\epsilon}_{ih} d\tilde{\epsilon}_{im} d\tilde{\epsilon}_{ic} \\ &= \frac{1}{(2\pi)^{3/2}} |\tilde{\boldsymbol{\Sigma}}|^{-1/2} \left(\prod_k \prod_t \frac{\lambda_{itk}^{y_{itk}}}{y_{itk}!} \right) \int_{\tilde{\epsilon}_{ic}} \int_{\tilde{\epsilon}_{im}} \int_{\tilde{\epsilon}_{ih}} g^{MV}(\tilde{\epsilon}_i)\; d\tilde{\epsilon}_{ih} d\tilde{\epsilon}_{im} d\tilde{\epsilon}_{ic}, \end{aligned}$$

where $g^{MV}(\tilde{\epsilon}_i) \equiv \left(\prod_k e^{\tilde{\epsilon}_{ik} \sum_t y_{itk}} e^{-e^{\tilde{\epsilon}_{ik}} \sum_t \lambda_{itk}} \right) e^{-\frac{1}{2}(\tilde{\epsilon}_i - \tilde{\mu})' \tilde{\boldsymbol{\Sigma}}^{-1} (\tilde{\epsilon}_i - \tilde{\mu})}$. Taken together, the posterior distribution is defined as

$$\begin{aligned} f(\tilde{\epsilon}_i|\mathbf{y}_i) &= \frac{f(\mathbf{y}_i|\tilde{\epsilon}_i) f(\tilde{\epsilon}_i)}{f(\mathbf{y}_i)} \\ &= \frac{g^{MV}(\tilde{\epsilon}_i)}{\int_{\tilde{\epsilon}_{ic}} \int_{\tilde{\epsilon}_{im}} \int_{\tilde{\epsilon}_{ih}} g^{MV}(\tilde{\epsilon}_i)\; d\tilde{\epsilon}_{ih} d\tilde{\epsilon}_{im} d\tilde{\epsilon}_{ic}}, \end{aligned}$$

and the expectation is

$$E^{MV}(\epsilon_i|\mathbf{y}_i) = \int_{\tilde{\epsilon}_{ic}} \int_{\tilde{\epsilon}_{im}} \int_{\tilde{\epsilon}_{ih}} \begin{bmatrix} e^{\tilde{\epsilon}_{ic}} \\ e^{\tilde{\epsilon}_{im}} \\ e^{\tilde{\epsilon}_{ih}} \end{bmatrix} f(\tilde{\epsilon}_i|\mathbf{y}_i)\; d\tilde{\epsilon}_{ih} d\tilde{\epsilon}_{im} d\tilde{\epsilon}_{ic}.$$

C.2 Derivation of $E^{UV}(\epsilon_{ik}|\mathbf{y}_{ik})$

The univariate expectation, $E^{UV}(\epsilon_{ik}|\mathbf{y}_{ik})$, is a special case of the multivariate expectation, $E^{MV}(\epsilon_i|\mathbf{y}_i)$. Replacing the conditional and marginal distribution functions with their univariate counterparts, the univariate posterior distribution is

$$\begin{aligned} f(\tilde{\epsilon}_{ik}|\mathbf{y}_{ik}) &= \frac{f(\mathbf{y}_{ik}|\tilde{\epsilon}_{ik}) f(\tilde{\epsilon}_{ik})}{f(\mathbf{y}_{ik})} \\ &= \frac{g^{UV}(\tilde{\epsilon}_{ik})}{\int_{\tilde{\epsilon}_{ik}} g^{UV}(\tilde{\epsilon}_i)\; d\tilde{\epsilon}_{ik}}, \end{aligned}$$

where $g^{UV}(\tilde{\epsilon}_{ik}) \equiv \left(e^{\tilde{\epsilon}_{ik} \sum\limits_t y_{itk}} e^{-e^{\tilde{\epsilon}_{ik}} \sum\limits_t \lambda_{itk}} \right) e^{-\frac{1}{2\tilde{\sigma}_k^2}(\tilde{\epsilon}_{ik}-\tilde{\mu}_k)'(\tilde{\epsilon}_{ik}-\tilde{\mu}_k)}$ and $k \in \{c, m, h\}$, and the univariate expectation is

$$E^{UV}(\epsilon_{ik}|\mathbf{y}_{ik}) = \int_{\tilde{\epsilon}_{ik}} e^{\tilde{\epsilon}_{ik}} f(\tilde{\epsilon}_{ik}|\mathbf{y}_{ik}) \, d\tilde{\epsilon}_{ik}$$

for $k \in \{c, m, h\}$.

D Appendix Tables

On the ensuing pages, we report Tables A1 through A8. Table A1 presents descriptive statistics for the tricoverage sample. Tables A2, A3, and A4 provide details on the claims, premiums, and deductibles, respectively, in the tricoverage sample. Tables A5 and A6 report the regression parameters estimates for auto and home, respectively. Table A7 reports the association parameter estimates for alternative samples A and B. Table A8 reports the least squares estimates of $\boldsymbol{\gamma}$ in the premium regression.

Table A1: Descriptive Statistics
Tricoverage Sample (294,917 household-years)

	Mean	Standard deviation	Minimum	Maximum
Auto:				
Driver 1 age (years)	56.10	14.70	19	99
Driver 1 female	0.33	0.47	0	1
Driver 1 single	0.22	0.41	0	1
Driver 1 married	0.63	0.48	0	1
Driver 1 insurance score	789.51	106.50	297	996
Driver 2	0.48	0.50	0	1
Driver 2 age (years)	50.28	12.93	16	94
Driver 2 female	0.91	0.28	0	1
Driver 3+	0.04	0.21	0	1
Young driver	0.01	0.10	0	1
Vehicle 1 age (years)	4.43	3.59	-1	46
Vehicle 1 personal use	0.47	0.50	0	1
Vehicle 1 passive restraint	0.99	0.10	0	1
Vehicle 1 anti-theft	0.57	0.49	0	1
Vehicle 1 anti-lock brakes	0.79	0.41	0	1
Vehicle 2	0.53	0.50	0	1
Vehicle 2 age (years)	5.94	5.53	-1	83
Vehicle 2 personal use	0.55	0.50	0	1
Vehicle 2 passive restraint	0.94	0.24	0	1
Vehicle 2 anti-theft	0.46	0.50	0	1
Vehicle 2 anti-lock brakes	0.70	0.46	0	1
Vehicle 3+	0.05	0.22	0	1
Home:				
Home age (years)	45.05	27.20	0	206
Insured value (thousands of dollars)	153.31	75.63	1	3,250
Farm or business	0.02	0.15	0	1
Primary residence	1.00	0.04	0	1
Owner occupied	0.98	0.14	0	1
Number of families	1.16	1.89	1	99
Masonry construction	0.07	0.25	0	1
Distance to fire hydrant (feet)	401.83	514.82	0	30,000
Alarm or other protection	0.95	0.22	0	1

Note: Insurance score is based on information contained in credit reports.

Table A2: Claims

Tricoverage Sample (294,917 household-years)

	Auto collision		Auto comprehensive		Home	
Count	Frequency	Percent	Frequency	Percent	Frequency	Percent
0	265,692	90.09	285,923	96.95	273,984	92.90
1	27,186	9.22	8,495	2.88	18,886	6.40
2	1,890	0.64	467	0.16	1,872	0.63
3	140	0.05	30	0.01	159	0.05
4	6	-	-	-	12	-
5	3	-	2	-	2	-
6					2	-

Note: Dash indicates less than 0.01 percent.

Table A3: Premiums
Tricoverage Sample (294,917 household-years)

	Auto collision	Auto comprehensive	Home
Mean	200	127	548
Standard deviation	104	70	309
Minimum	20	6	50
1st percentile	60	34	204
5th percentile	82	48	265
10th percentile	97	58	296
25th percentile	129	81	359
Median	178	113	466
75th percentile	243	157	638
90th percentile	327	210	891
95th percentile	393	250	1,110
99th percentile	560	358	1,683
Maximum	2,520	2,524	10,224

Note: Amounts in dollars.

Table A4: Deductibles
Tricoverage Sample (294,917 household-years)

	Auto collision		Auto comprehensive		Home	
Deductible	Frequency	Percent	Frequency	Percent	Frequency	Percent
$50	-	-	34,007	11.53	-	-
$100	7,846	2.66	18,502	6.27	11,577	3.93
$200	65,672	22.27	128,599	43.61	-	-
$250	51,644	17.51	31,556	10.70	197,100	66.83
$500	159,702	54.15	78,098	26.48	70,567	23.93
$1,000	10,053	3.41	4,155	1.41	14,537	4.93
$2,500	-	-	-	-	1,044	0.35
$5,000	-	-	-	-	92	0.03

Note: Dash indicates deductible option not available.

Table A5: Regression Parameter Estimates - Auto
Tricoverage Sample (294,917 household-years)

	Collision		Comprehensive	
	Estimate	Standard error	Estimate	Standard error
Intercept	-0.998 *	0.135	-2.675 *	0.248
Driver 1 age (years)	-0.011 *	0.004	0.039 *	0.008
Driver 1 age squared (hundreds of years)	0.013 *	0.003	-0.048 *	0.007
Driver 1 female	0.067 *	0.021	-0.084 *	0.041
Driver 1 married	0.048	0.025	0.125 *	0.046
Driver 1 separated, divorced, or widowed	0.000	0.023	0.058	0.045
Driver 1 insurance score (tens)	-0.018 *	0.001	-0.013 *	0.001
Has 2 drivers	0.063	0.123	-0.135	0.214
Has 3+ drivers	0.529 *	0.158	0.058	0.255
Young driver	0.020	0.049	0.019	0.082
Driver 2 age (years)	0.012 *	0.005	0.006	0.009
Driver 2 age squared (hundreds of years)	-0.013 *	0.005	-0.002	0.008
Driver 2 female	0.097 *	0.034	-0.064	0.060
Driver 2 married	-0.207 *	0.047	-0.121	0.087
Driver 2 separated, divorced, or widowed	0.088	0.164	0.000	0.302
Vehicle 1 age (years)	-0.012	0.005	-0.028 *	0.006
Vehicle 1 age squared (hundreds of years)	-0.015	0.044	0.143 *	0.036
Vehicle 1 personal use	-0.010	0.014	-0.034	0.025
Vehicle 1 passive restraint	-0.078	0.062	-0.114	0.102
Vehicle 1 anti-theft	0.011	0.015	0.018	0.027
Vehicle 1 anti-lock brakes	0.026	0.016	0.039	0.030
Has 2 vehicles	0.281 *	0.056	0.689 *	0.095
Has 3+ vehicles	0.293 *	0.107	0.930 *	0.156
Vehicle 2 age (years)	-0.023 *	0.003	-0.020 *	0.005
Vehicle 2 age squared (hundreds of years)	0.031 *	0.010	0.019	0.018
Vehicle 2 personal use	-0.019	0.015	-0.035	0.027
Vehicle 2 passive restraint	0.075	0.039	-0.033	0.062
Vehicle 2 anti-theft	0.029	0.018	0.009	0.033
Vehicle 2 anti-lock brakes	-0.003	0.019	-0.023	0.032
Year dummies	Yes		Yes	
Territory codes	Yes		Yes	

Notes: Insurance score is based on information contained in credit reports. Territory codes indicate rating territories, which are based on actuarial risk factors, such as traffic and weather patterns, population demographics, wildlife density, and the cost of goods and services.
* Significant at the 5 percent level.

Table A6: Regression Parameter Estimates - Home Tricoverage Sample (294,917 household-years)

	Estimate	Standard error
Intercept	-1.968 *	0.250
Insurance score (tens)	-0.018 *	0.001
Home age (years)	0.003 *	0.001
Home age squared (years)	0.000	0.000
Insured value (tens of thousands of dollars)	0.015 *	0.001
Farm or business	0.098 *	0.047
Primary residence	0.631 *	0.228
Owner occupied	0.121	0.077
Number of families	-0.011	0.007
Masonry construction	0.048	0.029
Distance to fire hydrant (feet)	0.001	0.001
Alarm or other protection	0.019	0.036
Year dummies	Yes	
Territory codes	Yes	
Protection classes	Yes	

Notes: Insurance score is based on information contained in credit reports. Territory codes indicate rating territories, which are based on actuarial risk factors, such as traffic and weather patterns, population demographics, wildlife density, and the cost of goods and services. Protection classes gauge the effectiveness of local fire protection and building codes.

* Significant at the 5 percent level.

Table A7: Association Parameter Estimates - Alternative Samples

	Tricoverage sample (62,425 households; 294,917 household-years)			Alternative sample A (8,731 households; 78,579 household-years)			Alternative sample B (203,731 households; 1,019,170 household-years)		
	Estimate	95 percent confidence interval		Estimate	95 percent confidence interval		Estimate	95 percent confidence interval	
Variances:									
Auto collision	0.107	0.065	0.149	0.114	0.049	0.180	0.093	0.070	0.116
Auto comprehensive	0.399	0.221	0.577	0.342	0.068	0.616	0.402	0.300	0.505
Home	0.405	0.383	0.428	0.401	0.260	0.541			
Covariances:									
Auto collision and auto comprehensive	0.137	0.101	0.173	0.123	0.064	0.182	0.131	0.112	0.151
Auto collision and home	0.061	0.022	0.099	0.121	0.081	0.161			
Auto comprehensive and home	0.225	0.179	0.271	0.209	0.135	0.282			
Correlations:									
Auto collision and auto comprehensive	0.663	0.399	0.926	0.622	0.195	1.049	0.680	0.522	0.838
Auto collision and home	0.293	0.099	0.486	0.564	0.298	0.830			
Auto comprehensive and home	0.559	0.389	0.729	0.563	0.247	0.880			

Notes: The tricoverage sample comprises an unbalanced panel of households who held all three coverages (auto collision, auto comprehensive, and home) in one or more years between 1998 and 2006. Alternative sample A comprises a balanced panel of households who held all three coverages (auto collision, auto comprehensive, and home). Alternative sample B comprises an unbalanced panel of households who held both auto coverages (collision and comprehensive).

Table A8: Premium Regression Parameter Estimates
Home Sample (62,425 policies)

	Estimate	Standard error
Intercept	0.1706 *	0.0056
Insurance score (tens)	0.0004 *	0.0000
Home age (years)	0.0000 *	0.0000
Farm or business	-0.0010 *	0.0003
Primary residence	-0.0474 *	0.0024
Owner occupied	-0.1280 *	0.0050
Number of families	-0.0004	0.0002
Masonry construction	-0.0172 *	0.0001
Distance to fire hydrant (feet)	0.0000 *	0.0000
Alarm or other protection	-0.0013 *	0.0002
Year dummies	Yes	
Territory codes	Yes	
Protection classes	Yes	
Adjusted R^2	0.8295	

Notes: Insurance score is based on information contained in credit reports. Territory codes indicate rating territories, which are based on actuarial risk factors, such as traffic and weather patterns, population demographics, wildlife density, and the cost of goods and services. Protection classes gauge the effectiveness of local fire protection and building codes.

* Significant at the 5 percent level.