README

Overview

The Matlab code in this replication package implements a Monte Carlo study to illustrate estimation methodologies developed in the paper "Econometrics of Insurance with Multidimensional Types" by Aryal, Perrigne, Vuong, and Xu (2024).

It first generates synthetic/simulated data on individuals, including their probability of accidents θ, their risk aversion a, their choice of insurance contract (t, dd), an exogenous variable Z, the number of accidents J, and their corresponding damages $D_1, \cdots, D_J$. Then, based on individuals' insurance choices $\chi \in \{1, 2\}$, and observed accidents J, the Monte Carlo study aims to recover (i) The distribution of the probability of accidents θ; (ii) The joint distribution of the probability of accidents θ and risk aversion a. The main file Main.m runs all of the code to generate the data for the 5 figures in the paper.

On a MacBook Pro with an Apple M3 chip, 16GB of memory, and running Mac OS Sonoma 14.7, the code took approximately 45 minutes to run on MATLAB 2024b.

Data Availability and Provenance Statements

☑ This paper does not involve analysis of external data (i.e., no data are used or the only data are generated by the authors via simulation in their code).

Computational requirements

Software Requirements

☑ The replication package contains one or more programs to install all dependencies and set up the necessary directory structure. [HIGHLY RECOMMENDED]

- The code was executed using MATLAB Release 2024b with the following toolboxes installed: Simulink, Econometrics Toolbox, Statistics and Machine Learning Toolbox, Optimization Toolbox, and Parallel Computing Toolbox.

Controlled Randomness

☑ Random seed is set at line 13 of the main program Main.m
☐ No Pseudo random generator is used in the analysis described here.

Memory, Runtime, Storage Requirements

Summary Approximate time needed to reproduce the analyses on a standard (CURRENT YEAR) desktop machine:

☐ <10 minutes
☑ 10-60 minutes
☐ 1-2 hours
☐ 2-8 hours
☐ 8-24 hours
☐ 1-3 days
☐ 3-14 days
☐ > 14 days

Approximate storage space needed:

☐ < 25 MBytes

☑ 25 MB - 250 MB

☐ 250 MB - 2 GB

☐ 2 GB - 25 GB

☐ 25 GB - 250 GB

☐ > 250 GB

☐ Not feasible to run on a desktop machine, as described below.

Details: The code was last run on a **MacBook Pro with an Apple M3 chip, 16GB of memory, running macOS Sonoma 14.7**. With approximately 200GB of free disk space, the full run took around 45 minutes to complete.

Description of programs/code

- Main.m will generate synthetic data and produce figures illustrating the estimation of the insurance model. The script does not require any user inputs, and the results, including visualizations, are automatically generated. Main.m will run all the other programs contained in this replication package.
- First, Main.m generates a random sample $\{(\chi_i, J_i, D_{1i}, \ldots, D_{J_i i}, Z_i), i = 1, \ldots, N\}$, where the sample size $N = 100,000$. In this stage, the function GDP.m is called within Main.m, and subsequently, contract_choice.m is invoked within GDP.m

    a. The probability of accidents θ and risk aversion a are marginally distributed as Beta(2,3) on $[0, 1]$ and 10^{-3}Beta(1,3) on $[0, 10^{-3}]$, respectively. A negative association between risk and risk aversion is introduced through a Gaussian copula with correlation $\rho = -0.5$.

Damages $D_1, \cdots, D_J$ are exponentially distributed with a mean of $5,000$, denoted by $H(\cdot)$, while the number of accidents J follows a Poisson distribution with parameter θ.

b. The exogenous variable Z follows a uniform distribution on $[100, 200]$ and is independent of $(\theta, a, J, D_1, \cdots, D_J)$.

c. Two insurance coverages are considered: $(t_1, dd_1) = (3.25Z, 1000)$ and $(t_2, dd_2) = (700, 500)$. The variable $\chi = 1$ if the insurance coverage (t_1, dd_1) is chosen; $\chi = 2$ otherwise.

d. In one simulation, a sample of $N = 100,000$ triplets (θ_i, a_i, Z_i) is drawn from the joint distribution $F(\theta, a)$ and $U(100, 200)$. The value of Z_i determines the pair of offered coverages and the frontier (5) in the paper. Individual i selects coverage $(3.25Z_i, 1000)$ (i.e., $\chi_i = 1$) if their risk $\theta_i \leq \theta(a_i, Z_i)$, and coverage $(700, 500)$ otherwise. Based on θ_i, the number of accidents J_i is drawn from a Poisson distribution with mean θ_i, and damages $D_{1i}, \ldots, D_{J_i i}$ are sampled from the damage distribution $H(\cdot)$.

e. Using one random sample, the estimation procedure (explained in the subsequent sections) is applied. This process is repeated $R = 100$ times to generate a 90% confidence interval for the conditional density $f_{a|\theta}(\cdot|\cdot)$.

- Second, Main.m implements a three-step estimation procedure: A detailed description of the three-step estimation procedure can be found in Section 4.1 of Aryal, Perrigne, Vuong, and Xu (2024).

(1) Estimate $f_{\theta|Z}(\cdot|\cdot)$ using a constrained Generalized Method of Moments (GMM) and kernel smoothing. The constrained GMM estimator is implemented by using $M = 4$ moments which is the integer part of $\log N/(\log \log N)$.

In this step, the functions moments_func.m, moments_var_func.m, and inverse_finite_moments.m are called within Main.m. Subsequently, lambda_i.m and shifted_legendre_poly.m are invoked within inverse_finite_moments.m. Moreover, moment_equation.m is called within lambda_i.m, and LegendreShiftPoly.m is called within shifted_legendre_poly.m.

(2) Estimate $f_{\theta|\chi,Z}(\cdot|1, \cdot)$ by adapting the approach from Step 1, but conditioning on $\chi = 1$. In this step, the "true" $f_{\theta|\chi,Z}(\cdot|1, \cdot)$ is also obtained through simulation to allow for comparison between estimates and the true distribution. In estimation, we apply the constrained GMM estimator with $M = 4$ moments on its support $[\theta(\overline{a}, z), \theta(0, z)]$. This step also requires estimates of the probability $\Pr(\chi = 1|Z)$ of choosing coverage 1 as well the kernel regression of $J_i(J_i - 1) \ldots (J_i - m + 1)$ on Z_i with $m = 1, \ldots, 4$. Kernel estimators are performed using rule-of-thumb bandwidths.

In this step, the functions density_condition.m, chipr_condition.m, and conditional_density.m are called within Main.m. In addition, inverse_finite_moments.m is invoked in density_condition.m and inverse_finite_moments_condition.m is invoked in conditional_density.m. Sequentially, lambda_i_condition.m and shifted_legendre_poly.m are invoked within inverse_finite_moments_condition.m

(3) Estimate $f_{a|\theta}(\cdot|\theta)$ by plugging in estimates of $\partial\theta(a,z)/\partial a$, $\partial\theta(a,z)/\partial z$, and $\partial \Pr(\chi = 1|\theta(a,z), z)/\partial z$ into the following expression:

$$f_{a|\theta}[a(\theta,z)|\theta] = -\frac{\partial\theta[a(\theta,z),z]/\partial a}{\partial\theta[a(\theta,z),z]/\partial z} \times \frac{\partial \Pr[\chi = 1|\theta,z]}{\partial z},$$

where $\frac{\partial \Pr(\chi=1|\theta,z)}{\partial z}$ is obtained through numerical differentiation, and the estimation of $\Pr[\chi = 1|\theta, Z]$ uses the following expression:

$$\Pr[\chi = 1|\theta, Z] = \frac{f_{\theta|\chi,Z}(\theta|1,Z)\Pr(\chi = 1|Z)}{f_{\theta|Z}(\theta|Z)}.$$

In this step, density_condition.m, chipr_condition.m, and conditional_density.m are invoked within Main.m to estimate $f(\theta|\chi = 1, Z)$, $\Pr(\chi = 1|Z)$, and $f(\theta|\chi = 1, Z)$, respectively. In addition, f_atheta.m is called within Main.m to compute the true density $f(a|\theta)$ for comparison

- Main.m generates five figures in the main body of the article.

  (i) Figure 2 displays the observations (θ_i, a_i) for one simulated sample. The frontiers $\theta(a,z)$ when z varies from 110 (right curve) to 190 (left curve) provide the locuses of points (θ, a) for which the z-individuals are indifferent between the two coverages (t_1, dd_1) and (t_2, dd_2)

  (ii) Figure 3 shows the histogram of the number of accidents for one simulated sample. To generate this figure, indifference_curve.m function is called within Main.m

  (iii) Figure 4 presents the estimated marginal density of the expected number of accidents θ. It displays both the true density and the 90% confidence interval, based on $R = 100$ repetitions.

  (iv) Figure 5 displays the estimated conditional density $\hat{f}_{a|\theta}(\cdot|0.4)$, conditional on $\theta = 0.4$, as θ is distributed as $B(2,3)$ with a mean of 0.4. The figure also shows the 90% confidence interval. The range of $a(0.4,z)$ is $[0, 10^{-3}]$ when z varies.

  (v) Figure 6 displays the estimated conditional density $\hat{f}_{a|\theta}(\cdot|0.6)$, conditional on $\theta = 0.6$, along with the 90% confidence interval. The range of $a(0.6,z)$ is $[0, 0.44 \times 10^{-3}]$ when z varies.

Instructions to Replicators

- Run `Main.m` to execute all steps in sequence. The program was last run in October 2024. It operates without requiring any user input, and the results, including visualizations (Figures 2 to 6), are generated automatically. `Main.m` calls all the other Matlab code in this replication package to generate synthetic data and performs the three-step estimation procedures, as described in detail above.

List of tables and programs

The provided code reproduces:

☐ All numbers provided in text in the paper
☐ All tables and figures in the paper
☑ Selected tables and figures in the paper, as explained and justified below.

Figure #	Program	Line # in Main.m	Output file	Note
Figure 1	n.a. (no data)			See Lemma 1
Figure 2	Main.m	39-56	figure2.png	
Figure 3	Main.m	60-81	figure3.png	
Figure 4	Main.m	480-504	figure4.png	
Figure 5	Main.m	649-676	figure5.png	
Figure 6	Main.m	678-707	figure6.png	

Contact Information

For questions or support regarding this code, please contact Haiqing Xu at h.xu@austin.utexas.edu