All Actuaries Summit
Think bigger, live better
1-3 May 2024 • Gold Coast

Actuaries Institute.

Optimal Illusion – A look at the practical limitations of price optimisation

Prepared by Hugh Miller and Tom Moulder

Presented to the Actuaries Institute
2024 All-Actuaries Summit
1-3 May 2024

Actuaries Institute
ABN 69 000 423 656
Level 2, 50 Carrington Street, Sydney NSW 2000
P +61 (0) 2 9239 6100 | actuaries.asn.au

Abstract

Price optimisation in insurance seeks to modify premiums away from risk cost in ways that maximise business objectives, such as increasing profit while achieving a target volume. Knowledge of customers' price sensitivities can lead to sophisticated approaches to optimise premiums.

While the practice naturally attracts controversy, it is generally accepted as effective. However, significant uncertainties and complexities arise on more careful analysis. Issues with model inaccuracy and complexities with multi-year impacts mean that establishing superior performance is less clear.

This paper builds on existing research by exploring the situations where optimisation approaches can potentially struggle:

- How do strategies play out over multiple years, and what is the impact on a portfolio's price sensitivity over time?
- How do nonlinearities in modelled demand curves, such as those induced by link functions, affect optimisation?
- What are the practical impacts of model uncertainty on performance?
- How do strategies fare in situations where customer price sensitivity can vary year to year?
- How important is random price testing to maintaining accurate demand models?

The paper looks at theoretical and simulated results, with an emphasis on understanding the robustness of standard optimisation approaches under plausible uncertainty assumptions on a multi-year basis.

1 Introduction

Price optimisation in insurance refers to pricing changes to achieve a business objective (such as profit or volume) that goes beyond standard risk pricing. If two cohorts with differing price sensitivity are identified, then a business will usually be better off (in the short term, at least) increasing prices slightly for the price insensitive group and decreasing for the other group.

Insurance, particularly personal line general insurance, offers unique possibilities for price optimisation given the ability to tailor premiums (individualised or small-cohort pricing), and the large amount of information collected on those individuals. Effective optimisation depends on detailed modelling of individuals' propensity to buy insurance against other variables, so is of interest to actuaries who have the technical skills and subject area expertise.

Despite the above, there is relatively little public research on optimisation, for two key reasons:

- Commercial sensitivities – Companies that employ price optimisation see it as an area of high commercial confidence, and in many cases a competitive advantage.
- Ethical concerns – Since optimisation involves increasing prices for some customers for reasons other than risk, there is moral ambiguity around its appropriateness.

Ethical concerns have led to regulation in some jurisdictions. For instance, many US states prohibit the use of factors for pricing outside of those used in risk rating, or optimisation that is unfairly discriminatory (see for instance the CAS Price Optimisation White Paper). Similarly the UK Financial Conduct Authority introduced regulations from 2022 to ban price walking – a form of optimisation trading off profit between new business and renewal customers (Financial Conduct Authority 2021). In Australia, the practice is not regulated, although it is sometimes raised as an industry concern either directly or indirectly[1].

---

[1] Complaints about insurer 'loyalty taxes' such as those raised by the NSW Insurance Monitor in 2018, will intersect with concerns on price optimisation.

There is an implicit assumption in the discussion above that price optimisation is effective; otherwise companies would not invest heavily in implementing optimisation. However, effectiveness has not been heavily studied, and this is the motivation for our paper. It turns out that:

- It is easy to optimise on variables that lead to perverse, or at least unhelpful, portfolio outcomes (section 3.1)
- Retention as well as elasticity has a role to play in optimising premiums, and that the adopted link function as a significant impact on resulting premiums (section 3.2)
- Overconfidence in the ability to estimate elasticity leads to degradation in optimisation performance in a way that is not always obvious (section 3.3)
- Assumptions around model structure have profound impacts on results, when no 'true' model structure is known in advance (section 3.4)
- Optimisation leads to a more elastic customer base over time, which can carry longer-term implications (section 3.5)
- The ability to estimate customer elasticity (a core element of optimisation) relies on a good experimentation setup and testing volumes – which carries a significant cost (sections 3.6 and 3.7).

These results call into question the overall viability of optimisation, or at least creates a need to carefully understand the risks on an insurer's customer base, and the inherent limitations of the process.

In some places our work builds on that of Semenovich & Petterson (2019), which showed theoretical underperformance of optimisation in simplified circumstances when there is uncertainty in elasticity estimation. Our work embeds this in a full customer model, as well as other aspects of model risk.

2 Simulated data setup

2.1 Baseline setup

Most of our research questions are challenging to answer with analytic results; simulated data becomes easier and more practical. Our aim is to set out a plausible base case with dynamics that might reflect a typical insurer context, but with a focus on elasticity considerations. This involves some significant simplifications elsewhere, including:

- A basic relationship between profit margins and premium.
- An optimisation at a fixed point in time (so all policies renew at the same time), which simplifies the practical issues around optimising in real time.
- A focus on renewal business only, rather than incorporating new business strategy.
- Ignoring multiline effects, where the value of customers holding multiple products is accounted for.
- No explicit modelling of competitor premium effects.
- Considering retention function structures that yield a smooth relationship between premium and retention. We do not consider retention functions that include steps or discontinuities.

While these are likely important and would in practice need to be modelled, they have a smaller impact on the types of issues explored in this paper.

Our baseline setup has the form set out below. We deliberately mimic a generalised linear model structure, using a logit link function, so that the inverse link is $g(x) = \frac{1}{1+e^{-x}}$:

- $n = 100{,}000$ customers, indexed by i
- Latent variables $x_{i1} \sim N(0, 0.2)$, $x_{i2} \sim N(0, 0.3)$, $x_{i3} \sim N(0, 0.3)$, $x_{i4} \sim N(0, 0.5)$ for generating differences between individuals and related parameters $\phi_1 = 0.1$, $\phi_2 = -0.3$

- Risk cost (including other expenses) R_i sampled from a gamma distribution with shape $\alpha = \left(\frac{\mu_{ri}}{\sigma_{ri}}\right)^2$ and rate $\beta = \frac{\mu_{ri}}{\sigma_{ri}{}^2}$, for an individual with mean μ_{ri} and standard deviation σ_{ri}, defined with constant coefficient of variation $\frac{\sigma_{ri}}{\mu_{ri}} = 0.4$. The mean for an individual is set to $E(R_i) = \exp(ln(800) + x_{i1} + x_{i2})$.
- An average retention rate of $\mu_d = 80\%$, with retention $d(x_i) = g(g^{-1}(\mu_d) + \phi_1 x_{i1} + x_{i3}) = g(z_i)$
- A standard profit margin of 15%, so that base premium equals $P_i = R_i/0.85$, and base profit is $\pi_i = P_i - R_i$.
- $r_i = R_i d(x_i) = R_i d_i$ is the expected risk cost allowing for retention.
- True elasticity parameter $e_i = -\big(3 + 2(\phi_2 x_{i1} + x_{i4})\big)$, subject to a maximum of -0.5. This is defined so that the demand function with respect to a percentage price change p_i is $d(x_i, p_i) = g(z_i + e_i p_i)$.
- Maximum price changes of $\pm 20\%$ in a given year.

Under this setup and notation profit can be expressed as $\pi_i = (P_i(1 + p_i) - R_i)g(z_i + e_i p_i)$.

2.2 Flexibility and variation in setup

The inclusion of the latent variables $\phi_1$and ϕ_2 allows for some relationships to be set across the simulation. We include a slight positive correlation between the risk model and base retention probability ($\phi_1 = 0.1$), and between the risk model and elasticity parameters ($\phi_2 = -0.3$).

The setup also gives flexibility to test alternative relationships, and add uncertainty to estimates of elasticity, as we do later in the paper.

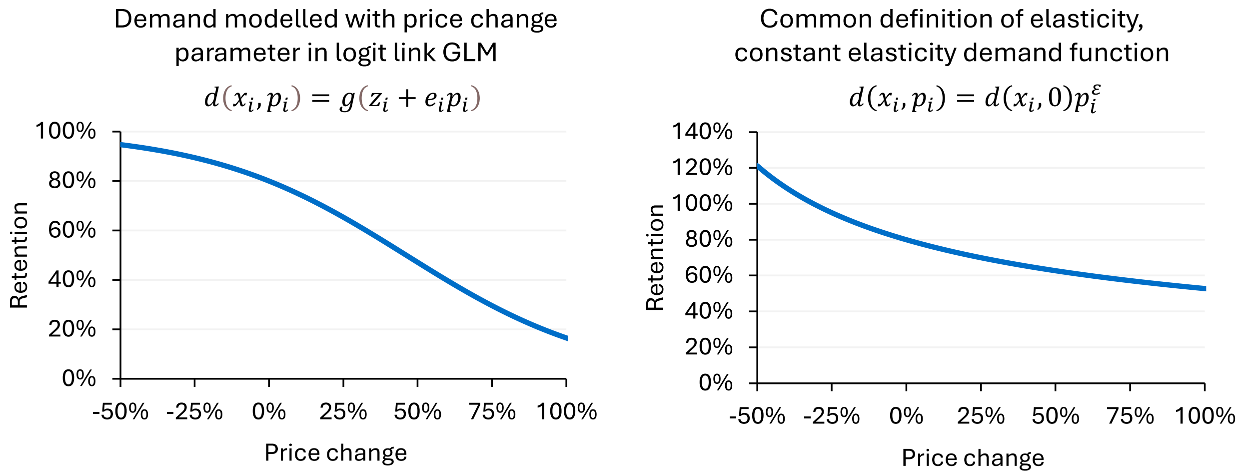
2.3 Defining the relationship between price change and demand (elasticity)

The common definition of price elasticity of demand is based on the percentage change in demand associated with a given percentage change in price. Under this definition, for some elasticity parameter ε_i, demand can be expressed as $d(x_i, p_i) = d(x_i,\ 0)p_i^{\varepsilon_i}$ (Talluri and Van Ryzin 2005).

The key awkwardness of this formulation is that individual-level retention probabilities are bound between zero and one, whereas the common formulation is unbounded. Applying a link function, such as the logit, to ensure retention lies in this allowable range is a natural solution, but it changes the relationship between elasticity and demand by imposing diminishing returns as the probability approaches one. This motivates our setup in section 2.2. The elasticity parameter is setup there as linear (prior to logit transformation) – this matches conventional approaches to modelling elasticity within the GLM framework. A comparison of demand as functions of these alternative formulations is illustrated below.

The **elasticity parameter** e_i defined in section 2.2 is therefore conceptually different from the standard definition of price elasticity of demand. The analyses in this paper relate to this elasticity parameter, rather than the price elasticity of demand ε_i. We note that under a constant elasticity parameter e_i, the implied price elasticity of demand ε_i will change at a policy's premium changes.

Figure 1 – Relationship between retention and price change under different approaches for capturing price elasticity.



2.4 Optimisation

Our main optimisation routine is designed to target a specific volume of written risk cost C_R (under expected retention rates), and maximise profits at that volume. Using the parameterisation from Section 2.1, this yields:

$$\text{maximise} \sum_{i=1}^{n} (P_i(1 + p_i) - R_i) d(x_i, p_i)$$
$$\text{subject to} \sum_{i=1}^{n} R_i d(x_i, p_i) = C_R$$

As shown in Semenovich & Petterson (2019), an optimality condition under this specification is that all customers must be priced either at their upper or lower bound, or must have identical $\frac{\partial \pi_i}{\partial r_i}$. This is intuitive, as if two customers have different $\frac{\partial \pi_i}{\partial r_i}$, then profit could be improved without changing written risk cost by moving price up for one policy, and down for the other.

Our main optimisation routine was designed based on this idea. Implementation was a combination of two phases:

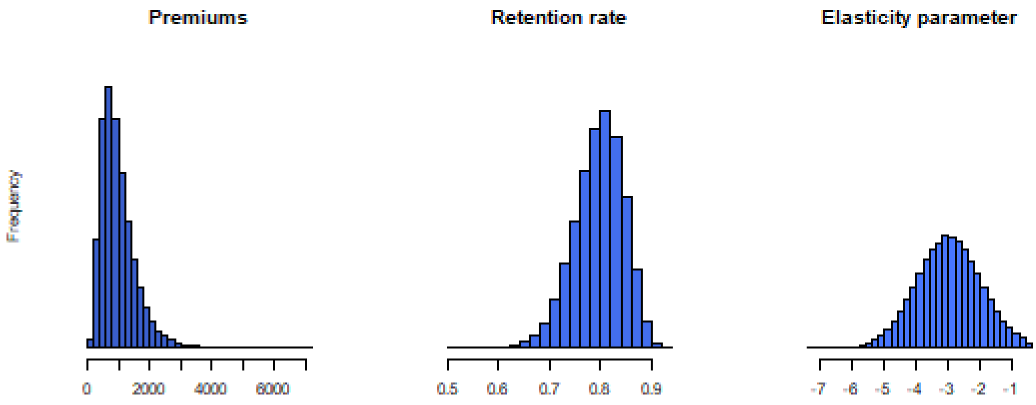
1. Newton steps to identify a single fixed price that that achieves the overall target volume
2. Optimisation steps based on calculated derivatives of profit against demand (risk-cost weighted), $\frac{\partial \pi_i}{\partial r_i}$. Observations with larger values see premium decreases and low values increases, until all observations are at price change limits or have an equal derivative.

Performance of the optimiser is relatively good – 100,000 observations in about three seconds.

2.5 Basic results

The distributions of base premiums and retention (before price changes) as well as the elasticity parameter are shown in Figure 2.

Figure 2 – Histograms of base premiums, retention rates and elasticity parameters



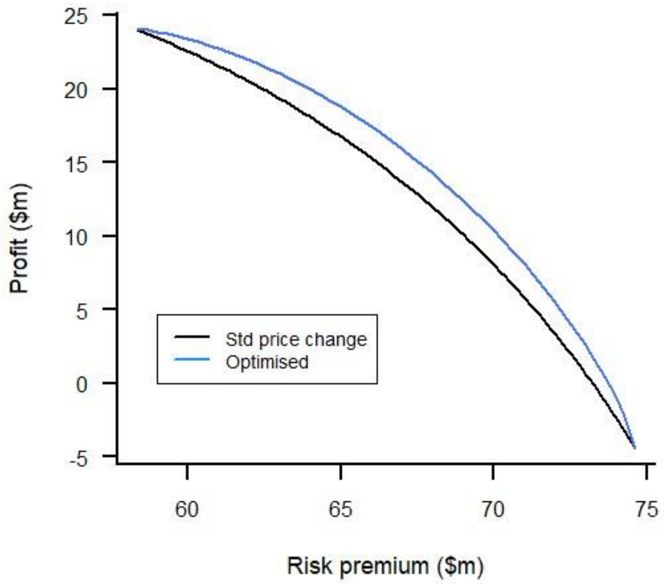
The dataset is structured in a way that overall elasticity is relatively low; increasing all premiums will increase (one-year) profits. The black curve in the figure below shows profit against risk cost volume under uniform price changes; if prices are increased one moves leftwards along the curve towards lower premium and more profit.

Gains from price optimisation are moderate. Targeting a risk volume of $68m, profit of $11.96m under a uniform price change rises to $14.25m under optimisation, a 19% lift in profit, equalling about 3% of risk cost. At the extremes of the curves all price changes are at the minimum/maximum and so standard and optimised results are the same.

Table 1 – Basic optimisation results

	Customer volumes	Avg premium	GWP ($m)	Risk premium ($m)	Profit ($m)
All price change 0%	79,578	1,004.49	79.98	67.98	12.00
Uniform price change	79,600	1,004.04	79.96	68.00	11.96
Optimised prices	79,712	1,036.20	82.25	68.00	14.25

Figure 3 – Standard and optimised curves trading off risk volumes and profit



Price changes will correlate with other customer dimensions. Resulting price changes for an optimisation targeting $68m risk cost are shown below. Figure 4 shows the overall distribution of price changes – around 20% of policies are at the pricing extremes, with the remainder distributed within. In Figure 4 far more policies are assigned the maximum price change of +20% than at the minimum of -20%. This is a function of the selected target risk premium. Selecting a lower target risk premium would result in a significantly larger proportion of policies at the lower bound.

Elasticity is an important driver of optimised premium changes, but not the only one. Figure 5 shows that those with higher (less negative) elasticity parameters tend to have higher premium increases. Figure 6 shows that those with higher initial retention rates are also assigned higher price changes. Finally, Figure 7 shows that there is not a pattern by base premium; this is deliberate in our setup to focus on risk cost rather than absolute customer numbers.

Figure 4 – Distribution of price changes, basic optimisation

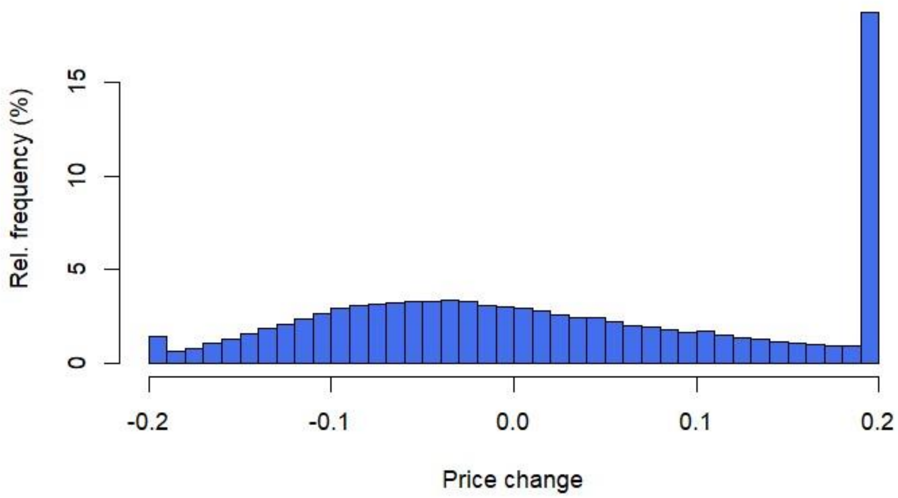


Figure 5 – Price change by elasticity parameter

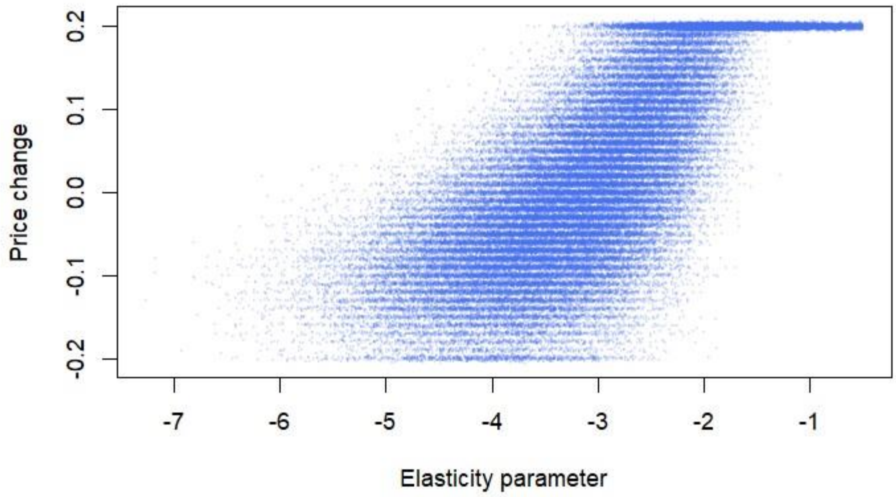


Figure 6 – Price change by initial retention rate

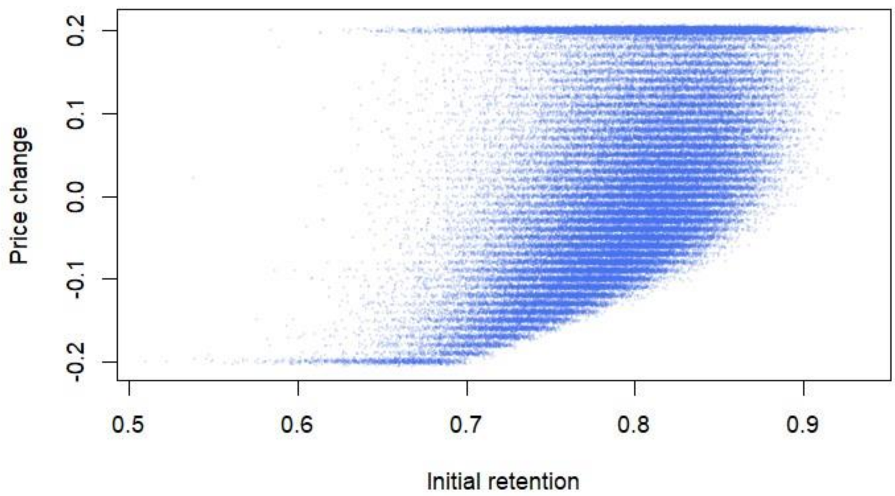
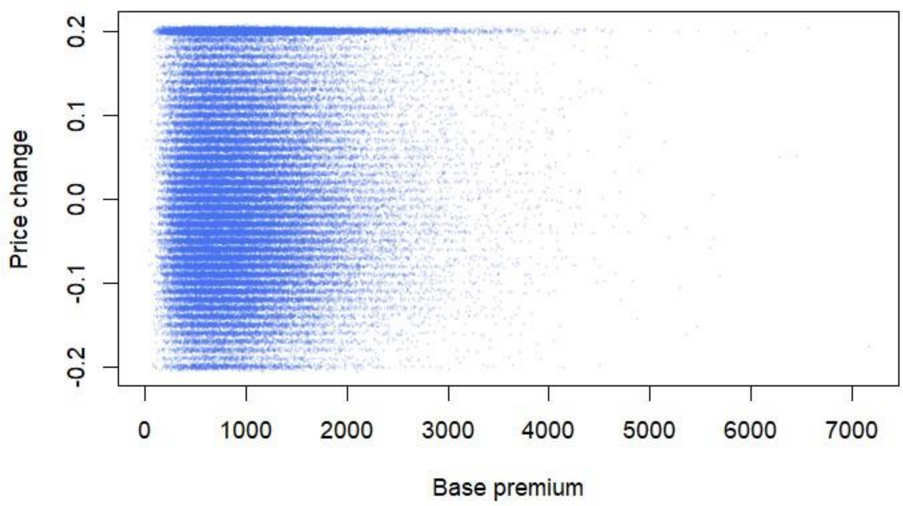


Figure 7 – Price change by initial premium



2.6 Role of optimisation versus headline strategy

One important aside, before deeply optimisation-specific results, is the relative size of optimisation-based improvements compared to headline pricing. The potential gains are material in our example, but modest compared to the impact of aggregate changes.

This is shown in the curves Figure 3 (which indexes price changes from -20% through to 20%), where the 'gap' between standard and optimised curves is only a small fraction of the overall movement of the curve. To use the example of Table 1, a similar increase in profit to perfect optimisation can be achieved with a uniform 3% increase in prices (with a corresponding 2% reduction in expected risk premium written).

While we do not claim that the simulated setup is truly representative of the market, it does suggest two implications:

- Reasonable time should be invested into headline pricing decisions and strategic positioning – not just optimisation.
- When considering the potential risks and inaccuracies of optimisation work (as we do in this paper) the comparison to simple price changes can be instructive. This includes situations where optimised profits are up, but no more than if a simpler price increase had been put through.

3 Results

3.1 What is the right thing to optimise?

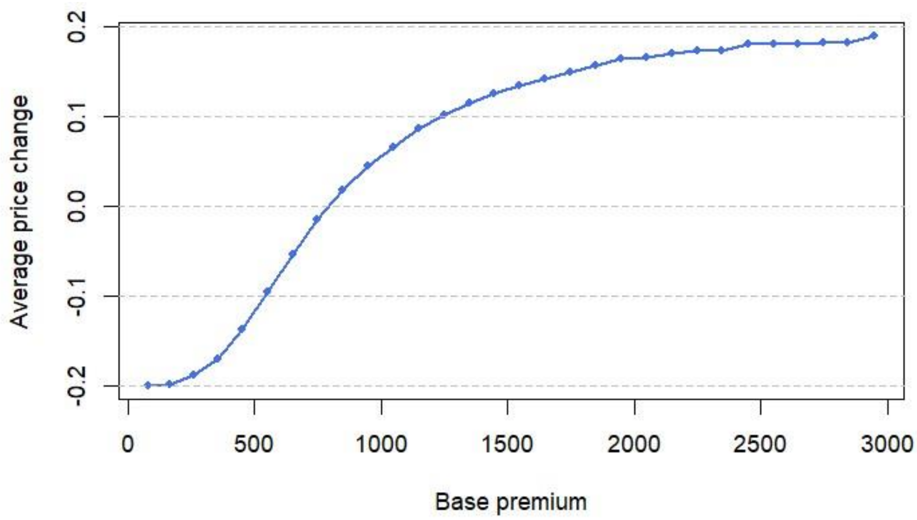
Given full knowledge of the system, it is possible to vary prices to improve profit, but there remain choices as to how we define a good optimisation. Our selection above is on risk premium volume, but the choice is not automatic. A pure profit maximisation exercise (that is, set all premiums to maximise profit), under our setup, will move all prices to the maximum, since we have price elasticity that are smaller than demand. However, this approach is short-term and better aligns with global strategy rather than the more complex optimisations we are seeking.

A plausible alternative is maximising profit with respect to a level of customer retention numbers. Customer volumes is a standard (and objective) metric, so seems natural. However, it produces perverse impacts on written premium. Price reductions are concentrated on people with low base premiums (as they are 'cheaper' to buy numbers) and increases focused on large premiums (as increases generate larger absolute profit gains).

Figure 8 shows average price changes by base premium if we change our scenario to optimise for base premium rather than written risk cost. The results are very different, with premium increases concentrated in large policies, rather than the even spread of Figure 7. This can be viewed as a strategy that seeks to lose significant risk premium volumes subject to a customer numbers constraint.

Compared to an optimisation on risk volumes that delivers similar customer numbers, optimising on customer numbers results in 3% lower risk premium and 14% higher profit.

Figure 8 –Price changes by starting premium obtained when optimising for customer numbers rather than risk premium volumes



While such a strategy is legitimate if customer numbers is genuinely a key metric, forcing away larger premiums will ultimately shrink an insurer's written premium relatively quickly. For this reason we view risk premium volume as the correct way to view the portfolio since it provides an internally consistent focus on profit margins. Customers with similar profit margins but different premiums are treated similarly.

An additional option is to optimise for profit with respect to a target gross written premium. This should work reasonably similarly to risk premium, except that since gross written premium varies directly with price (whereas risk premium does not), it adds some extra dynamics that require control. To take an extreme example, if all policies had reverse elasticity (demand increases with price increase), an optimisation on gross written premium would fail to take advantage.

3.2 Role of elasticity and retention in determining optimised price changes

The figures in Section 2.5 illustrated the twin role of elasticity and retention rate in determining price changes. We can examine this by considering the derivative of profit against written risk cost $\partial\pi/\partial r$ (i dropped for convenience). This is the key determinant of whether a customer's premium will be increased or decreased by optimisation.

$$\frac{\partial\pi}{\partial r} = \frac{\frac{\partial\pi}{\partial p}}{\frac{\partial r}{\partial p}} = \frac{P(1+p)-R}{R} + \frac{P}{eR(1-d(x))}$$

The second term in this expression includes both (negative) elasticity parameter e and retention rate $d(x)$. All else equal:

- Customers with more negative values of e (more elastic) will have a less negative $\partial\pi/\partial r$ (meaning less profit is sacrificed to achieve a unit increase in demand) and are therefore more likely to see premium decreases under optimisation.
- Customers with higher initial retention $d(x)$ will have more negative values of $\partial\pi/\partial r$, and be more likely to see premium increases. That is, they appear to be less elastic.

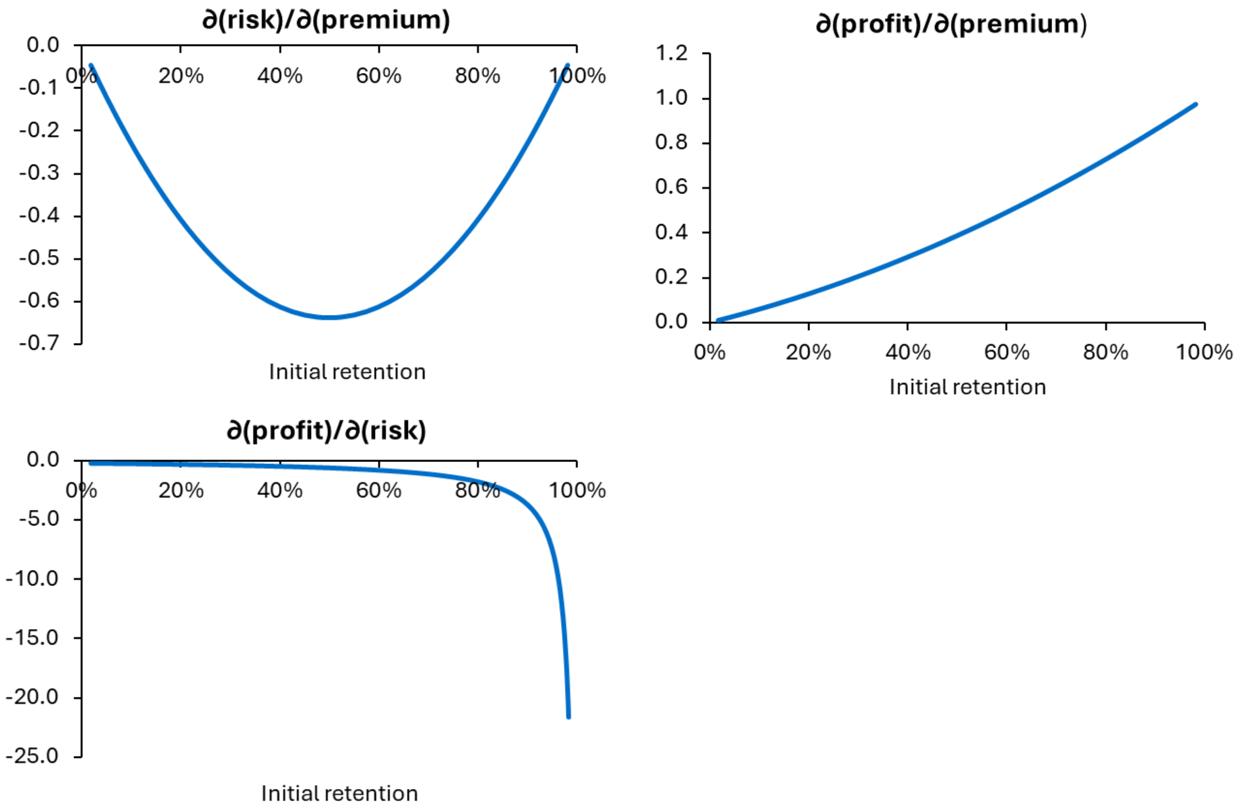
This influence of retention on apparent elasticity occurs partly due to the dampening effect of the inverse logit link function.

Figure 9 illustrates this retention effect by considering how each of the derivatives in the expression above changes based on a customer's initial retention. We see that written risk changes most quickly with premium at both high and low initial retention due to the inverse logit link. Profit increases more quickly with premium when initial retention is high. Combined, this means that $\partial\pi/\partial r$ decreases rapidly as initial retention increases. The hyperbolic shape of the relationship between $\partial\pi/\partial r$ means that customers with high modelled base retention are likely to be assigned substantial premium increases in optimisation.

The relationships shown in Figure 9 (and therefore optimised premiums) are strongly influenced by the adopted link function in the demand model. In practice, an inverse logit link is usually selected with little consideration, however, it should be noted that this makes an implicit assumption that customers with retention nearer to 50% are more elastic than those at the extremes. Section 3.4 explores this further.

While outside the scope of this paper, we also note that risk cost also appears in the expression for $\partial\pi/\partial r$. This means that errors in risk cost estimation will also deteriorate optimisation performance, particularly for customers where $d(x)$ is low, such that the first term in the expression above carries greater weight.

Figure 9 – Key derivatives for optimisation as base demand varies(a)



(a) Show for an example customer with e_i = -3, the average value in our simulated data setup.

Figure 10 approximates the proportion of the variation in $\partial\pi/\partial r$ that comes from variation in initial retention as opposed to variation in the elasticity parameter, under our standard setup with logit link structure. While not a perfect measure, this can be thought of as a proxy for the relative influence of variation in elasticity parameter and initial retention on optimisation gains.

We do this under simplified version of our baseline setup, where customers' elasticity parameter e_i and initial retention parameter z_i are independent. We simulate datasets with a range of different standard deviations of both e_i and z_i, and for each calculate the standard deviation of $\partial\pi/\partial r$ at base premiums. We approximate the proportion of variation in $\partial\pi/\partial r$ that is due to initial retention z_i by comparing the change in standard deviation of $\partial\pi/\partial r$ if the variation in e_i was set to 0, versus if the variation in z_i was set to zero.

The baseline data described in simulated has a standard deviation of e_i of approximately 1, and standard deviation of z_i of approximately 0.3. As such, we might expect around 28% of optimisation gains to be related to variation in initial retention rather than the elasticity parameter. We see that as variation in the elasticity parameter decreases, we quickly reach a point where over half of optimisation gains are due to variation in initial retention.

Figure 10 – Approximate proportion of variation in $\partial\pi/\partial r$ due to variation in initial retention, for data with average initial retention of 80%.

Standard deviation of elasticity e_i	Standard deviation of initial retention parameter z_i: 0	0.1	0.2	0.3	0.4	0.5	0.6
0.0		100%	100%	100%	100%	100%	100%
0.2	0%	54%	71%	79%	84%	88%	90%
0.4	0%	36%	53%	64%	71%	77%	81%
0.6	0%	25%	40%	51%	59%	66%	72%
0.8	0%	16%	28%	38%	46%	53%	60%
1.0	0%	11%	20%	28%	35%	42%	49%
1.2	0%	8%	15%	21%	28%	34%	40%
1.4	0%	6%	12%	18%	23%	29%	35%
1.6	0%	5%	11%	16%	21%	26%	31%

While not the focus of this paper, we tested a similar analysis on a simulated data designed to look more like a new business portfolio. That is, with low "retention" rate, and higher elasticity. In this case the role of initial retention variation was substantially lower than shown in Figure 10 for our baseline setup.

3.3 Revisiting the costs of elasticity misestimation

Semenovich & Petterson (2019) show a 'fool's gold' effect in optimisation contexts, showing that misestimation in both the estimation of elasticity and the measurement of optimisation's effect can lead to significant gaps between expected and true optimisation performance. Under simplified assumptions (most notably a linear setup and very small price adjustments only) a key result is that the profit improvement can be expressed as a decay due to error, which we label T:

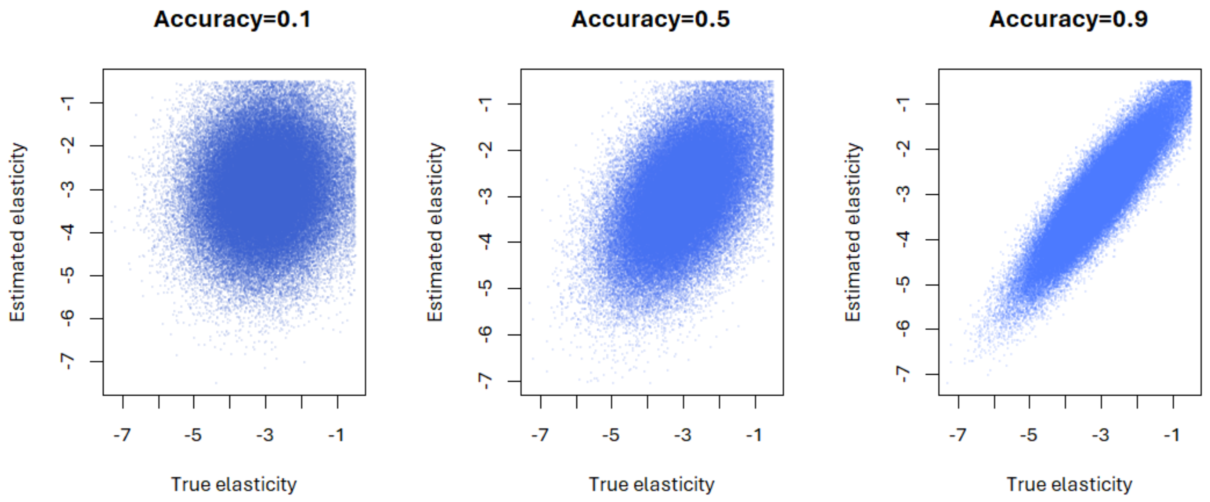
$$T = \frac{Actual\ profit\ improvement}{Expected\ profit\ improvement} = \frac{\sigma_b}{\sqrt{\sigma_a^2 + \sigma_b^2}}$$

Where σ_b is the spread of the derivative of profit with respect to demand ($\partial\pi/\partial d$) and $\sigma_a$is the spread of error in the estimation of this quantity. This quantity reflects the correlation between the original and noisy estimate of the derivative.

Under simulation, we can use a similar approach to add noise to simulated elasticity parameter e, to reflect a less than perfect parameter estimate. We do this by holding the overall spread of the elasticity parameter constant but increasing the relative component of additional noise to achieve a desired correlation T. We refer to this value as accuracy throughout this section.

We can then see how optimisation performance decays when optimising on the noisy parameter but measuring performance on the true parameter. Examples of different levels of noise are shown in Figure 11.

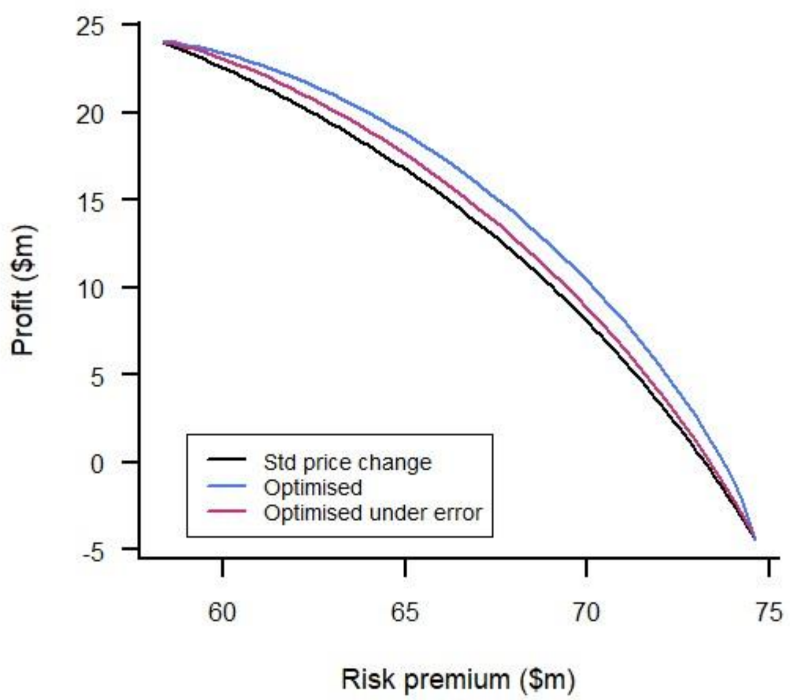
Figure 11 – True and noisy elasticity parameters under different levels of accuracy T



Under our standard setup and accuracy of 0.5, we can measure performance and construct the optimisation frontier, shown in Figure 12. Perhaps surprisingly, the frontier, rather than being midway between the standard and optimal curves, sits closer to the standard, implying less gain under optimisation. Despite this, if evaluating optimisation performance using the noisy elasticity parameter estimate, **expected** gains are similar to the true optimal curve.

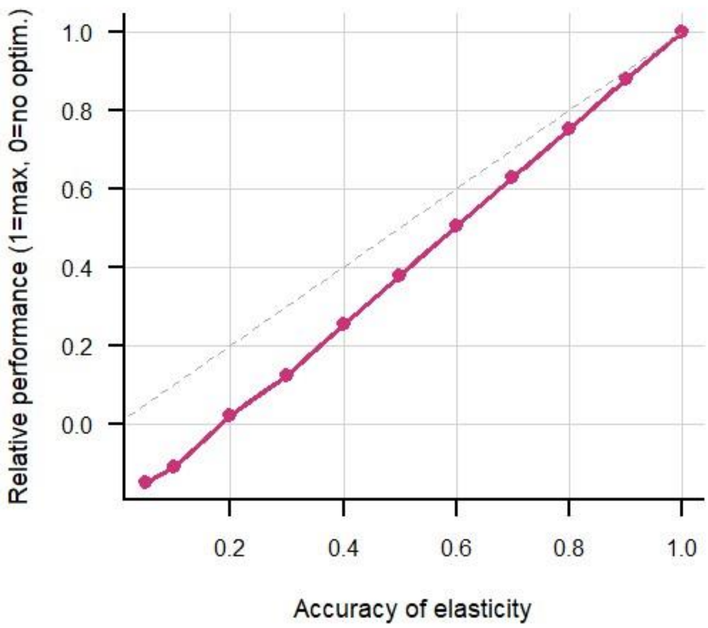
Importantly, we note that most of the deterioration is seen in risk volumes rather than profit. Since optimisation puts through an overall increase in premiums, this still increases profits (as per the standard price change curve), but risk volumes come in below expected. For example, at the target of $68m risk volume, the profit increase is still 17% (compared to 19% for the true curve), but actual volumes are $67m.

Figure 12 – Standard, optimal and optimal subject to error (T=0.5) frontiers



We can characterise this performance as a percentage of the distance between the two curves (along a vertical line); at that level of actual risk volume, what fraction of the way between standard frontier and optimal is achieved. For $T = 0.5$, the result is about 40%. As we vary the accuracy, this performance appear to move linearly, but with a slope greater than one, as shown in Figure 13 below. We note that the definition of T is based on error in the estimate of the elasticity parameter, which differs from the definition in Semenovich & Petterson (2019).

Figure 13 – Performance of optimisation under noise – 0% is standard curve and 100% is optimal curve



The figure shows that in this setup, accuracy below 0.2 actually sees **worse performance than undertaking no optimisation at all**. We believe this is a new result – that optimisation can be damaging if there is a poor hold on elasticity.

The negative performance is largely a function of the nonlinear function $g(x)$ that links price change to retention. When retention is above 50% the function is concave. This means that, given two identical policies, randomly moving one up and the other down will lower average retention rate (under Jensen's inequality). The degree of impact can vary with the position on the curve and size of price movement. But as an example, if two policies start at 80% retention and price changes move by 0.3 on the linear predictor (so adjusted retention is 84.3% and 74.8%), then the average retention drops to 79.5%.

Importantly, the slope of the performance decay varies depending on the setup. If elasticity is large and variable relative to retention, the decay with inaccuracy will be large. Figure 14 shows a much stronger decay when the elasticity parameter is doubled and retention variability halved. Indeed in this case the accuracy must be above 0.5 before any benefit is seen.

Conversely, in situations where the elasticity variation is small, and retention variability large, the slope of decay softens. Figure 15 shows the result with the elasticity parameter halved and doubled retention variation – in this case the optimisation carries benefits even when the elasticity model is entirely noise (0% accuracy). Here the model is effectively 'optimising on retention', which can still deliver some value; 50% of the optimisation gains are still retained.

Figure 14 – Performance of optimisation under noise – double the variation in elasticity parameter and halved for retention

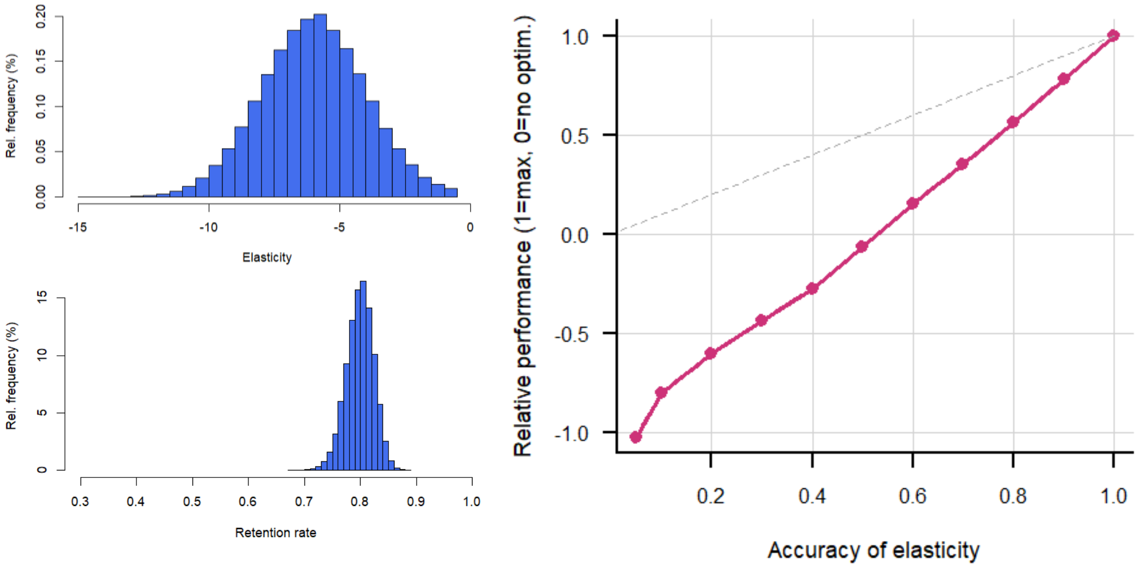
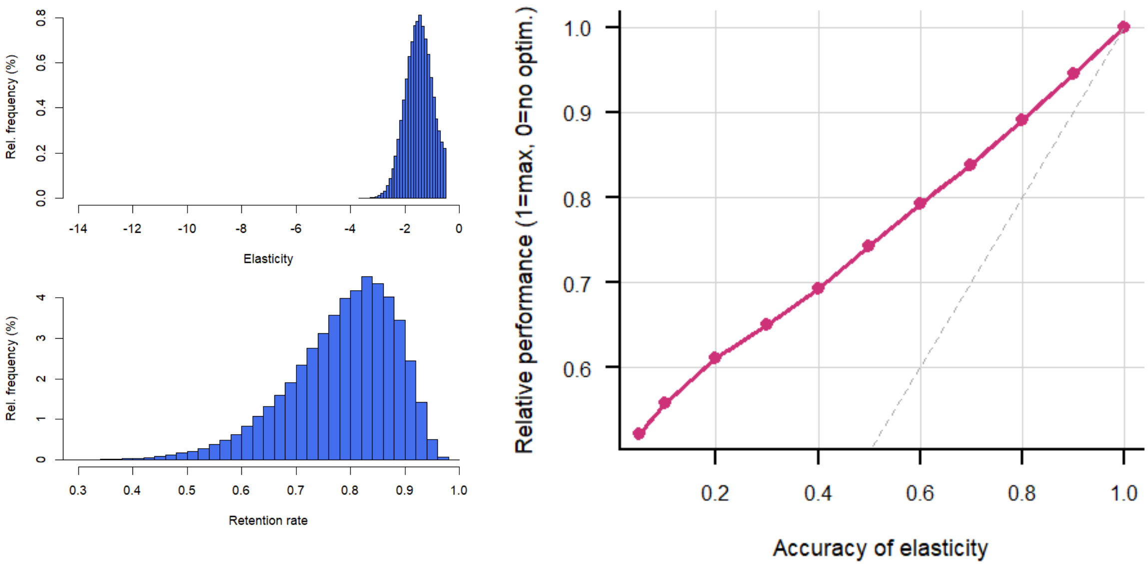


Figure 15 – Performance of optimisation under noise – half the variation in elasticity parameter and double for retention



As argued in Semenovich & Petterson (2019), the main protection against these effects (and understanding which of the above scenarios is closest to reality) is proper validation of the effectiveness of optimisation – either by constructing a suitable estimate from existing data, or setting up an independent validation group.

3.4 The criticality of assumed model structure

The underlying model structure connecting retention and elasticity is often set for practical and tractability reasons. Our setup in section 2.1 deliberately mimics a GLM structure with logistic link, since this is a practical way to estimate elasticity as a function of other terms in an internally consistent way from retention data. However, there is no intrinsic reason why this is the 'right' structure – no iron law saying that the logistic link is how elasticity terms and demand should be connected.

In a content of abundant data, the decision is less critical – enough interactions and nonlinear terms can be added to manhandle a sensible elasticity effect (Miller & Moulder 2018). However, price sensitivity testing is expensive and messy – data is not abundant. This means that the assumed structure is potentially very important.

We illustrate this by considering the following alternative structures:

- Alternative 1 – Elasticity effect also depends on retention linear predictor

$$g(x_i) = g\big(z_i + e_i^* p_i (z_i + 1)\big)$$

- Alternative 2 – Elasticity effect uses a probit link rather than logistic

$$g(x_i) = \Phi(z_i + e_i^* p_i)$$

- Alternative 3 – Elasticity is on a 'per dollar' basis rather than a percentage basis

$$g(x_i) = g(z_i + e_i^* p_i P_i)$$

In each case we generate elasticity under the alternative as truth, in a way that the sensitivity (as measured by a 5% price increase) is equivalent to our base setup. We then choose the optimal estimate of a misspecified elasticity, where we assume that the model form is our original $g(z_i + e_i p_i)$, but choose the best possible linear transform of the true elasticity. This means that our misspecified estimate correctly orders elasticity of all customers and gets elasticity effects right on average, but there is a degree of error induced by the misspecification of the retention function.

For each alternative we can optimise our model correctly and under misspecification, and then see the impact. The figures below show the results of this analysis. Figure 16 compares optimal price changes if the elasticity was known perfectly, against optimised price changes under the misspecified models. Figure 15 compares the optimised profit and risk premium achieved. We observe:

- **Alternative 1 - Extra retention correlation**: Price changes under the misspecified model have a similar general trend to the perfect model, but there is considerable variation at a customer level depending on their starting retention. On average, premium decreases are somewhat moderated in the misspecified model. We see slightly smaller premium decreases for policies with large price decreases under the optimal model, and vis versa. There are many examples of policies with over 20% difference in premium between the two models. Despite this variation, the optimisation gains under the misspecified models are not significantly less than under the optimised one. This demonstrates how slight model structure changes can have considerable impacts on customer level premiums, even if top level results change only minimally.
- **Alternative 2 – Probit link**: The probit link model suggests smaller price changes for customers that are pushed to the upper and lower bounds under the misspecified logit link model. This occurs as the probit link function results in modelled retention approaching asymptotes more rapidly with price adjustments. The misspecification results in small reductions in top line optimisation gains.
- **Alternative 3 – Per dollar elasticity:** Applying elasticity on a 'per dollar' basis has a significant impact on optimised premiums. There is now a strong incentive to price low risk cost policies at the upper bound, as the impact of increasing their premiums is relatively low. This change also results in apparent optimisation gains that are significantly larger than in other scenarios. While this scenario is likely not realistic, some weaker per dollar effects might be present in customer behaviour and it serves to illustrates how differences in true elasticity structure can cause large differences in resulting premiums and apparent optimisation gains.

Figure 16 – Price changes – optimal and optimal under misspecification for the three alternative setups

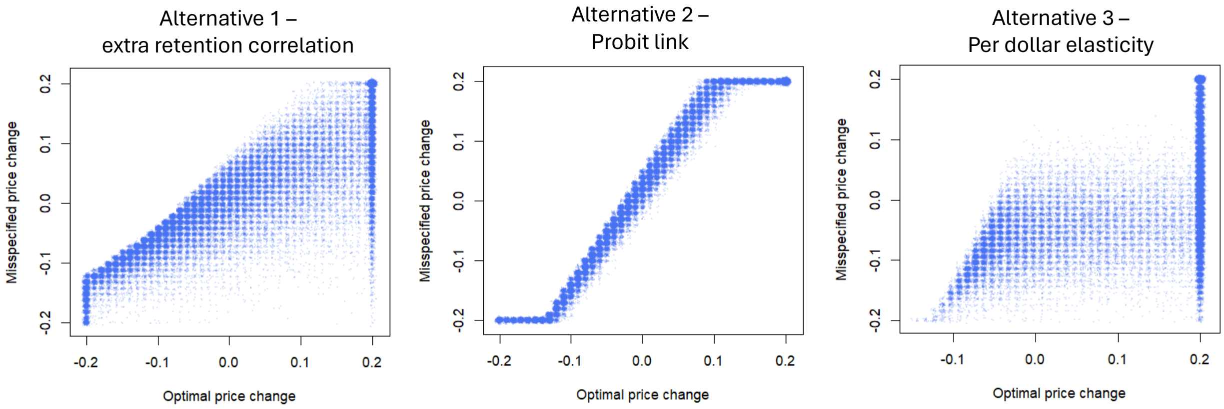
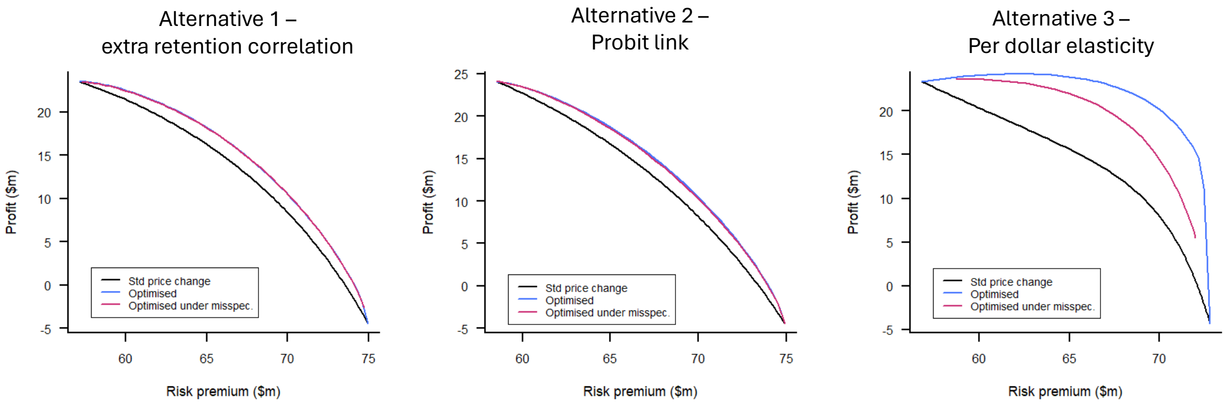


Figure 17 – Optimisation frontiers – optimal and optimal under misspecification for the three alternative setups



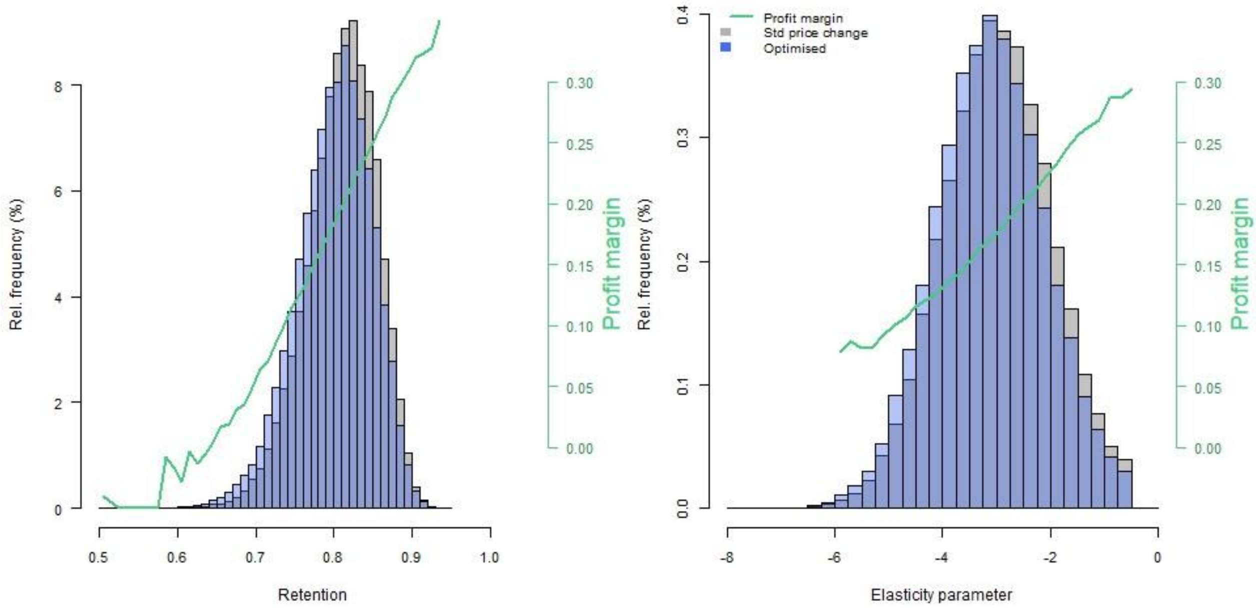
3.5 What is happening to the customer base over time?

Understanding the impact of price optimisation on the customer base over time is a natural question.

The core of optimisation is to increase prices for less elastic customers and decrease for more elastic. This necessarily implies a shift in towards a more elastic customer base than would otherwise be the case. We first use a simplified multiyear strategy to illustrate, where we optimise over a single year and adopt those prices unchanged over five years to see the impact.

Figure 18 shows the change in the distribution of elasticity parameter and base retention rates for non-optimised and optimised prices. In both cases the shift to more elastic and higher churn is clear – albeit on a five-year timescale that would allow for new business to partially balance. There is a sense that optimisation creates a more challenging future customer base, so the multiyear view becomes important. This is particularly true of profitability – since optimised profit margins are largest for customers with higher retention and lower elasticity; this is also shown on the figure.

Figure 18 – Histograms showing change in customer base after five years of standard pricing and (simplified) optimisation. Green curve shows targeted profit margins at different level of elasticity and retention.

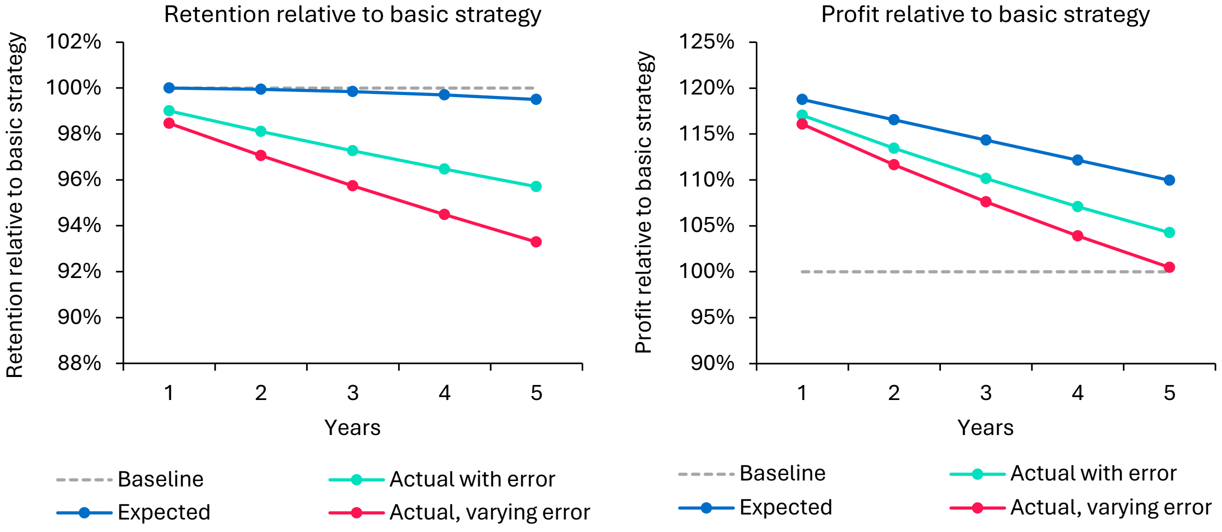


The effect of a more challenging customer base over time narrows the gap between optimised and simple pricing strategies for more distant years. The blue curve in Figure 19 shows that the initial 19% boost in profit (relative to uniform pricing) decays to just 10% by year five; the ability to generate extra profits dissipates as the fraction of high retention and lower elasticity customers dissipate.

The figure shows two other decay curves:

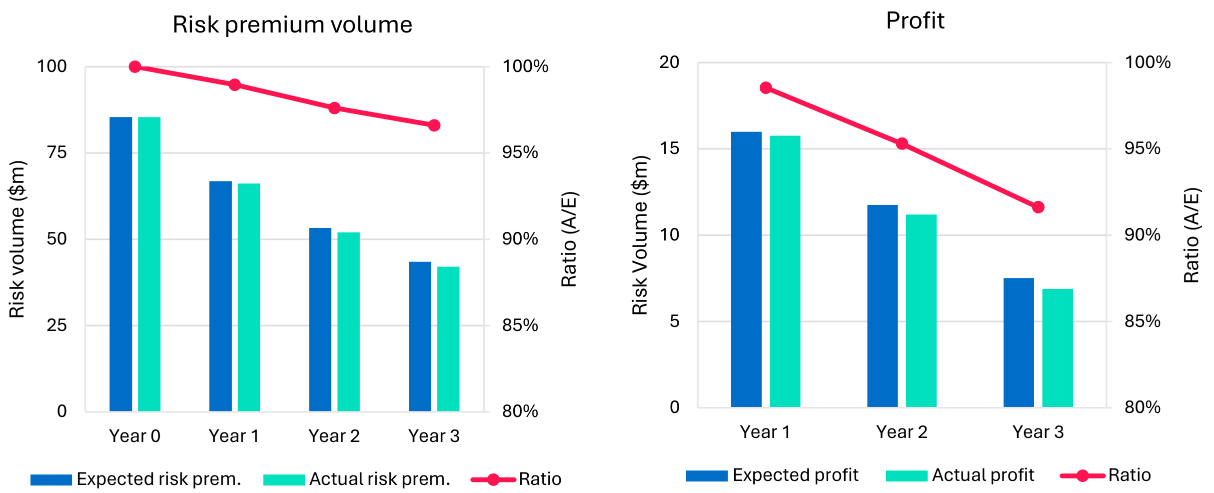
- With elasticity misestimation (50% accuracy, as per Section 3.2) the decay is more rapid as the falling customer load accelerates the profit decay. Total custom numbers also decay consistently relative to uniform pricing under elasticity misestimation.
- With elasticity misestimation and **varying elasticity** (so estimated elasticity is unchanged, but the 'true' elasticity parameter is resampled each quarter around this), we see slightly faster decay again, in both profit and retention. By year five profit is the same as under uniform pricing, and retention is 7% lower. Combined, this reflects performance that is substantially worse than uniform pricing. This sort of varying error is potentially an additional source of profit leakage to track in pricing validation work.

Figure 19 – Retention and profit relativities compared to simple pricing strategies



Finally, we note that although we have provided results on a simplified multiyear basis, results do tend to hold under different arrangements. For example, we ran a multiyear optimisation of our base setup, where the target was maximum cumulative profit subject to a target risk premium retained in the final (third) year. As shown in Figure 20, the gap between expected and actual risk premium grows steadily over the time period. However, the drop in profit is sharper and accelerates after the first year; the lost volume in the first year contributes to a snowballing effect on profits in later years too.

Figure 20 – Example impact of elasticity misestimation for a three-year optimisation



There is a more general question around how properly implemented multiyear strategies should vary prices relative to one-year optimisations. Generally, a longer-term view should place more value on retaining customers and so temper price increases. A full treatment of this is beyond the scope of this paper, but it is worth noting that as with section 3.1, the specifics around setup can drive different behaviours. In the example above, the optimal strategy over three years is to push up prices a little more initially, and the make up for it in later years with a softer pricing strategy to balance out and achieve the desired risk premium volume at year 3.

3.6 Challenges of modelling elasticity

Elasticity modelling as a causal problem

Estimating elasticity is a causal modelling problem. We are concerned with the treatment-effect of changing price on renewal probability. As a result, it comes with the standard challenges of casual modelling, including unobserved counterfactuals and concerns around confounding.

Moreover, elasticity modelling for optimisation is a heterogenous treatment effect estimation problem. We are not only interested in average elasticity, but also differences in elasticity between customers. This is more challenging again.

Model structure

There is an evolving literature on machine learning approaches for heterogeneous treatment effect modelling, with techniques such as causal forests as in Athey et al. (2018) as a prominent example. While the application of these methods to insurance price elasticity is an interesting area with early indications of promise, for example in Guelman & Guillén (2014) and Verschuren (2022), the tools to run these models are relatively immature, and to our knowledge have not been widely deployed by Australian insurers.

In practice, elasticity is often modelled as part of standard retention modelling. This is commonly in the form of a GLM (or regularised equivalents) that includes price change terms in the model. Interactions between price change terms and other parameters are used to capture differences in elasticity between different cohorts of customers. We limit our attention to this GLM model structure in this paper.

We note that while tree-based machine learning models like XGBoost and Random Forests are commonly used by insurers, they are not suitable for demand modelling in their raw form. This is because optimisation relies on smooth estimates of the relationship between price and retention, which tree-based methods to not produce by default.

The need for random price flexes in historical data

As a causal modelling problem, estimating elasticity requires a degree of historical price exploration. If a model is to estimate elasticity for different cohorts of customers, then that price exploration must occur for all subsets of customers.

Ideally, a fraction of policies would have random price changes applied to their premium, so that they could be used for elasticity modelling in the future. However, there can be several barriers to doing this, including:

- Technological costs and challenges in implementing random price flexes
- Ethical or regulatory concerns
- Lost profitability by taking policies that would have otherwise been optimised, and instead assigning them less profitable random price changes.

In the absence of truly random price changes, elasticity models can be constructed by exploiting price variation due to historical price changes. This has significant limitations relative to systematic random price flexing. Historical price changes will generally be applied differently across different types of customers, and are fixed at a single point in time, which means that additional care is needed to allow for the possibly confounding influence of changes in customer behaviour over time.

Risks of training elasticity models on historically optimised data

Elasticity may also be modelled by exploiting variation in premium loadings due to historical price optimisation. That is, training retention models on historically optimised policies. This requires great care, as historical price changes are not randomly assigned.

While optimisation is likely to produce a wide range of price changes across the portfolio, any individual cohort of customers may have very little price exploration. For example, a highly elastic cohort of customers may have always received price decreases. This challenges the positivity assumption (Hernan & Robins (2020)) of causal modelling, which requires variation in treatment assignment for all sub-groups, and can lead to material degradation in model performance if not properly controlled for.

To demonstrate this, we compare the results of optimisations run under the following scenarios.

- **Perfect model** – Optimisation with full knowledge of customers' elasticity
- **Retention model trained on random price flexed data –** Training data includes random price flexes between -20% and +20%.
- **Retention model trained on optimised data –** Training data is taken from the outputs of the basic optimisation described in Section 2.5.

All retention models are trained as binomial GLMs using an inverse logit link function, and contain main effects for x_{i1} to x_{i4}, plus a price change term to model elasticity. Note that in this simple setup we use only a single price change term that applies to all customers, to reflect a scenario where the retention model does not properly control for variation in historically optimised prices.

Table 2 shows the parameter estimates from these each of these models. We also show the "true parameters", which are the implied coefficients from the simulated demand relationship as specified in Section 2.1. For these true parameters, the intercept and elasticity parameter e are based on average values across the customers in the simulated data.

The model that is trained on data with random price flexes accurately recreates the true parameters from the historical data. However, the model trained in optimised customer data substantially understates average elasticity. This occurs because of the correlation between other modelled variables and historical optimised price changes. For example the x_3 coefficient is lower in the model trained on optimised data, as high retention policies are generally given lower optimised price changes. This parameter captures some signal that would otherwise have been captured in the elasticity term.

Table 2 – True parameters and estimates from retention models

Parameter	Intercept	x_1	x_2	x_3	x_4	e
True parameters	1.38 (average)	0.10	0.00	1.00	0.00	-3.00 (average)
Price flex data model	1.38	0.11	0.00	1.00	-0.04	-3.00
Optimised data model	1.44	0.04	0.00	0.75	0.11	-1.84

Table 3 shows the results of optimisation under each of these models, as well as a uniform price change.

- While the price flex data model has lower profit than the perfect model due to fitting only a simple constant elasticity effect, it still offers a 7% improvement in profit over a uniform price change.
  - While not the focus of this paper, we tested a similar analysis on a simulated data designed to look more like a new business portfolio. Here, a main effect only model provided almost no benefit over a uniform price change strategy. This is in line with the observations in Section 3.2.
- The model trained on optimised data offers no incremental profit improvement over the uniform price change strategy, due to the biases in model parameter estimates. While offering no improvement in profit, optimising using this model does produce other undesirable outcomes. Notably:
  - A wider distribution of premium changes than the price flex data model, as the lower average elasticity estimate pushes more policies to the extremes.
  - Writing a level of overall risk premium that is different to that targeted in the optimisation run. In this example, running the optimisation to target a risk premium of $68m resulting in an actual risk premium of only $66m (-3%), as the model underestimated the loss of business from an increase in average premiums. The results in Table 3 have been adjusted to use a target risk premium that achieves an actual result of $68m, to provide comparability with other scenarios.

Table 3 – Optimisation results when optimising with retention models

Scenario	Customer volumes	Avg premium	GWP ($m)	Risk premium ($m)	Profit ($m)
Uniform price change	79,600	1,004	79.96	68.00	11.96
True parameters	79,712	1,036	82.25	68.00	14.25
Price flex data model	79,637	1,014	80.84	68.00	12.84
Optimised data model	79,615	1,004	79.96	68.00	11.96

We note that the deterioration in performance seen here is an extreme example, as the main effect only elasticity parameterisation means that the model is not able to control for the biases in the optimised training data. At the other extreme, if we train a retention model that includes all interactions between the elasticity term and relevant other variables, the optimisation performs well, even if trained on optimised data.

Nonetheless, in practice is it challenging to perfectly parameterise a model to control for all biases in historical price changes. This example serves to demonstrate how optimisation performance can deteriorate if historical biases are not fully controlled for.

3.7 Volumes of data needed to estimate elasticity interactions

Even if training a retention model on data containing random historical price flexes, significant amounts of data can be required to accurately estimate elasticity for different subgroups of customers.

This is because estimating interactions (and variation in elasticity is estimated as such a model interaction) requires substantially more data than estimating main effects. For example, Gelman et al. (2020) shows that under reasonable assumptions, it would require 16 times as much data to estimate an interaction that is half the size of a main effect.

The exact amount of data required to accurately estimate interactions depends on several factors, including the size of the sub-population for which the interaction is being estimated, and the size of the interaction itself.

The magnitude of historical price variations is also a key influence on required sample sizes in elasticity modelling. Large historical price variations result in larger swings in demand, and more accurate estimates of elasticity – this is conceptually analogous to modelling a large treatment effect. Correspondingly, small historical price changes are analogous to smaller treatment effects, and are significantly harder to model.

As standard errors increase with the square root of sample size, halving the size of price changes in historical data (for example, moving from ±20% to ±10%) could be expected to increase the required size of the modelling dataset by a factor of four.

This is a consideration when designing price flexing processes. Larger price changes may be less desirable from a customer experience perspective, but can also substantially reduce the number of customers to which random price changes need to be applied. We note that this is not the only consideration in selecting the size of price flexes. For example, insurers may which to collect data across the range of price flexes that may be considered in optimisation.

We use simulated data to examine the relationship between optimisation gains, the number of price-flexed training samples, and the magnitude of historical prices flexes. We simulate data follows:

1. Simulate a data under the baseline setup described in Section 2.1
2. Randomly select price changes between the specified upper and lower limit
3. Calculate expected retention under these selected price changes, and simulate binary retention outcomes
4. Randomly down sample the dataset to achieve the desired dataset size.

We then train a retention model on the simulated dataset, and use model to optimise premiums on the full simulated dataset of 100,000 customers. Retention models are built either as:

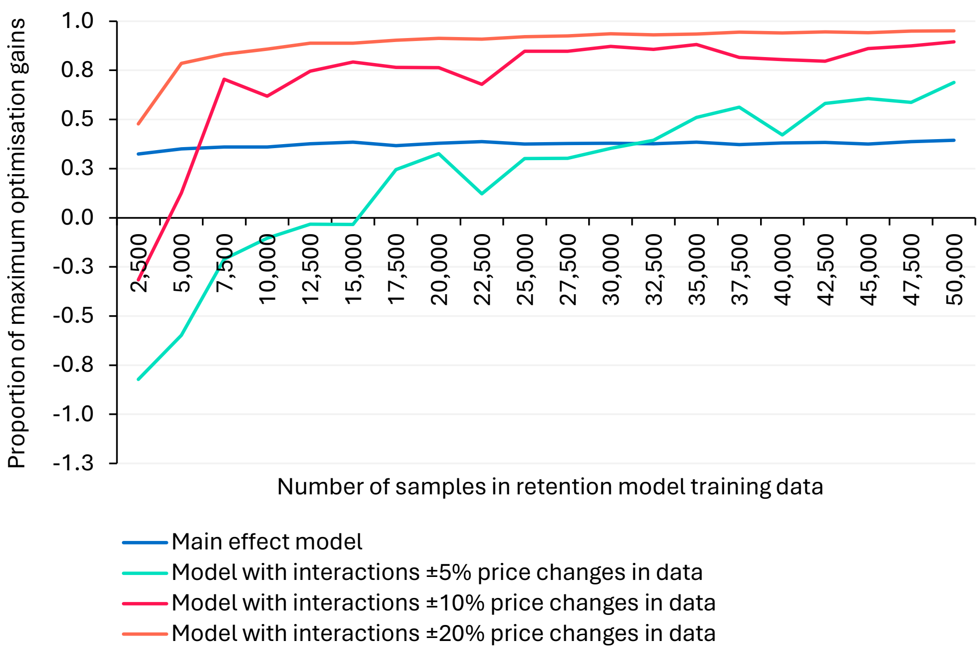
- A main effect model, using a single price change (elasticity parameter) term for all customers
- A model with interactions between the price change term and customer characteristics. We do this by ordering and grouping the true elasticity parameter e_i into five equally sized groups, and including a GLM model term for the interaction between the price change term and each group. This allows the model to fit a different elasticity parameter for each of the five groups. We note that this is somewhat optimistic, as we have given the model knowledge of 5 groups of customers with perfectly ordered true elasticities. In practice, required interactions would be less clear-cut, as true elasticity is unknown.

We repeat this process for a range of combinations of sample size and price change magnitude. Finally, we repeat the process 50 times for each combination, and average results across these repetitions. This smooths out the significant volatility that exists in performance for individual runs.

Figure 21 shows the results of this analysis. We plot the percentage of the maximum possible optimisation gains that is achieved by each model. We observe that:

- Where sample sizes and historical price changes are small, including interactions results in worse performance than a simple main-effect model. In some cases, performance can be worse than a simple uniform price change.
- Optimisation gains increase with sample size, but there are diminishing returns on sample size increases. Optimisation gains improve roughly in proportion to the square root of sample size. This may be expected as standard errors of parameters decrease with the square root of sample size, and Section 3.3 shows optimisation performance varying approximately linearly with elasticity parameter estimation accuracy.
- Optimisation gains at a given sample size increase significantly with the magnitude of historical price flexes. As above, models trained on data with a ±10% price flex would require approximately 40,000 observations (under a smoothed curve) to achieve the same performance as a model built on only 10,000 observations of data with a ±20% price flex.

Figure 21 – Relationship between retention modelling data size and optimisation performance



The exact number of samples required to satisfactorily model elasticity interactions will vary based on details of those interactions. Despite this, this simulation study suggests that sample sizes of at least 5,000 to 10,000 with random price flexes would be required to satisfactorily model simple interactions under one plausible setting. In practice, lower data quality, smaller interaction effect sizes, or complex or uncertain interactions may mean that even larger volumes are needed.

This may be tractable for major products sold by large insurers. However, it may present a significant limitation to optimisation for small insurers or lower volumes products. This is compounded by the preference for recent data, and desire to only randomly price flex a small proportion of the overall customer base.

4 Discussion

This paper demonstrates that while price optimisation can offer an opportunity for increasing profits, there are several complexities that may mean that benefits are significantly lower than estimated.

These complexities, while technically interesting, mean that care should be taken in both the implementation of optimisation, and measurement of its performance. In some situations, the potential costs and risks of price optimisation could call into question its overall value.

We summarise key lessons from our analysis below.

- Estimating the performance of optimisation using the same retention (and risk) models that are used for performing optimisation leads to a material overstatement of optimisation gains. This is the key theoretical result from Semenovich & Petterson (2019), which we observe in our simulated results. This highlights the importance of validating optimisation performance.
- The choice of what to optimise is very important, with different selections resulting in substantially different premium outcomes. For example, we see that optimising to a target number of policies written can result in significant biases in premium loadings by policy size. Large policies are given substantial premium increases, while policy volumes are maintained by writing a larger number of small policies at low (or negative) profit margins - likely an undesirable outcome in practice.

  We instead propose optimising to a target volume of risk premium as a preferable optimisation structure. Many other optimisation structures are also possible. Each will result in different pricing behaviour, and so should be considered carefully before deployment.

  We note that in practice, it may be desirable to optimise to multiple constraints. For example, maximising profit while constraining both written risk and policy volumes. While we do not explore this multi-constraint setting in this paper, we expect similar considerations to be relevant to those identified for the single-constraint setting. And the risk premium constraint remains the most relevant for managing a portfolio in an internally consistent way.
- Both elasticity and customers' initial retention have an impact on their optimised premiums, with the impact of initial retention influenced by the selected link function used in modelling.
- Error in estimation of elasticity significantly decays optimisation performance. Where accuracy of estimation is particularly low, this includes the possibility of worse performance than no optimisation at all. The impact of elasticity estimation is highly dependent on the relationship between retention variability and elasticity variability. If variation in elasticity is low relative to variation in retention, good optimisation gains are possible even with poor estimates of the elasticity parameter. Overfitting and adding noise to elasticity estimates will typically result in higher estimated optimisation performance, even though, as shown here, true performance will be lower.
- The structure of models used for estimating elasticity is very important. A correctly structured model will reflect the true relationship between price and retention, but this is not observed directly so is difficult to test. Where the true price retention relationship does not follow assumed structure, optimised premiums can end being materially differently to where they would have if the true structure was known. For example, changing retention to vary by dollar rather than percentage price changes results in significantly different premium changes and optimisation performance. In practice, the true relationship may be somewhere between dollar and percentage changes and include various kinks or steps in the elasticity curve.

We also see that in some cases overall optimisation performance may not be significantly impacted by model structure changes, even if the distribution of customer level premiums is significantly different.

- Optimisation leads to a more elastic and higher-churn customer base over time, as more elastic policies are assigned lower premiums loadings than less elastic ones. This change in customer base narrows the gap between optimised and fixed price strategies over time, such that is some cases there may be limited benefit from optimisation after several years – that is, most benefit is realised in the first years.
- Elasticity modelling is a causal estimation problem, and as such requires historical training data that includes a degree of price exploration. Ideally, this would take the form of a proportion of customers that are assigned a random price flex, although implement this may not be feasible in practice. An alternative is to train elasticity models on data from historically optimised policies. However, this can significantly deteriorate the quality of elasticity models and resulting optimisation, due to biases in which customers are assigned price changes in the models' training data. This problem can be partly addressed by correctly controlling for all confounding factors in the training data, although this is likely to be challenging in practice.
- Relatively large volumes of random price-flex data are required for elasticity modelling, as optimisation benefits from modelling interactions between elasticity and customer characteristics. These interactions require substantially more data than estimate main effects. For example, 5,000-10,000 observations with ±10% price flexes were required to achieve meaningful optimisation gains in our simulated setting. Additionally, while smaller price flexes may be more palatable for insurers, models trained with smaller price flexed require much more training data for the same performance. Halving the size of prices flexes increases required data by a factor of approximately four. The requirement for large volumes of price-flexed training data may present a barrier to optimisation for smaller insurers, or for smaller products.

Future work

There are several natural extensions of the analysis in this paper, which would work towards better understanding the limitations of price optimisation. These include:

- Extending the analysis to use actual customer data, rather than simulated results.
- Extending the analysis to relax the simplifications identified in Section 2.1. This includes reviewing results in a new business setting, considering multi-product effects, allowing for mid-term cancellations, and allowing for more complex retention functions that may include competitor effects or other discontinuities.
- Analysing the effectiveness and properties of techniques for validating optimisation performance in an unbiased matter. Examples include comparisons with non-optimised cohorts of the customer base, and inverse-propensity based methods as explored in Semenovich & Petterson (2021).

5 References

- Casualty Actuarial and Statistical Task Force (2015). *Price Optimization White Paper*. Accessed at: https://content.naic.org/sites/default/files/inline-files/committees_c_catf_related_price_optimization_white_paper.pdf, March 2024
- Talluri, K. & Van Ryzin, G. (2005) *The Theory and Practice of Revenue Management*. Springer, New York. Chapter 7.3.3.
- Financial Conduct Authority (2021) *General insurance pricing practices market study, Feedback to CP20/19 and final rules*. Accessed at: https://www.fca.org.uk/publication/policy/ps21-5.pdf
- Clarkson, A. & Black, A. (2016). *Price Optimisation: A case study in new pricing techniques*. Presentation to the IFOA and accessed at: https://www.actuaries.org.uk/system/files/field/document/G2%20ClarksonBlack.pdf

- Miller, H. & Moulder, T. (2018) *The missing link: squeezing extra performance out of predictive models*. Presented at the Presented to the 2018 Actuaries Institute General Insurance Seminar https://actuaries.logicaldoc.cloud/download-ticket?ticketId=4e6d3da5-569a-44d9-9c99-70895f4d4fa8
- Semenovich, D. & Petterson, J. (2019). *Unbiased Estimation of Economic Value of Pricing Strategies*. GIRO, Institute and Faculty of Actuaries. Accessed at: https://www.actuaries.org.uk/system/files/field/document/F7_Dimitri%20Semenovich_James%20Petterson_Paper.pdf .
- Semenovich, D. & Petterson, J. (2021). *Operational Decision Optimisation in Business Context*. Actuaries Summit 2021. Accessed at: https://actuaries.logicaldoc.cloud/download-ticket?ticketId=3c05b8ec-f7bb-49ed-9624-748ec224d745
- Guelman, L. & Guillén, M. (2014). *A causal inference approach to measure price elasticity in Automobile Insurance*. Expert Systems with Applications, Volume 41, Issue 2, Pages 387-396.
- Verschuren, R. (2022) *Customer Price Sensitivities in Competitive Automobile Insurance Markets*. Accessed at: https://arxiv.org/pdf/2101.08551.pdf
- Athey, S., Tibshirani, J. & Wager, S. (2018). *Generalized Random Forests*. Accessed at: https://arxiv.org/abs/1610.01271v4
- Hernan, MA. & Robins, JM (2020). *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC. Accessed at: https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/
- Gelman, A., Hill, J. & Vehtari, A. (2020) *Regression and Other Stores*. Cambridge: Cambridge University Press (Analytical Methods for Social Research). Chapter 16.4. Accessed at: https://avehtari.github.io/ROS-Examples/