

# Benchmarks for the benchmark approach to valuing long-term insurance liabilities: comment on Fergusson & Platen (2023)

Daniel Bauer 

University of Wisconsin-Madison, Madison, WI 53706, USA  
E-mail: [Daniel.bauer@wisc.edu](mailto:Daniel.bauer@wisc.edu)

(Received 20 February 2023; accepted 24 February 2023)

## Abstract

This article comments on the paper “Less-expensive long-term annuities linked to mortality, cash and equity” by Kevin Fergusson and Eckard Platen, appearing in this issue of the *Annals of Actuarial Science*. It adds two perspectives to their thought-provoking contribution. The first is a similarity to some recent work in quantitative finance on “deep hedging” that leverages machine learning models to find the cheapest replication strategy for a derivative payoff in a largely model-free setting. The second perspective engages with some of the interesting implications of their approach and draws parallels to literature in asset pricing and macro-finance. These perspectives point to the potential need for more fundamental shifts than the authors of the paper are advertising.

**Keywords:** Valuation; Derivative pricing; Financial economics; Long-term liabilities

## 1. Introduction

Fergusson & Platen (2023; FP23 in what follows) present a thought-provoking contribution on the valuation of long-termed insurance products such as life annuities. The tenet of their argument is that conventional market-consistent valuation principles arrive at values that are much too expensive because the payoffs can be “produced” – that is, replicated – at much lower costs. The contribution relies on and “aims to popularise” the benchmark approach to finance (Platen, 2006; Platen & Heath, 2006; among others).

This note adds two (related) perspectives to their contribution. First, I engage with the question on how these production costs are cast. As FP23 note, in popular financial models such as Black-Scholes, the approach proposed here will arrive at equivalent values. So, what are the alternative models where a difference emerges, and are they clearly superior? In this context, I connect to recent work that is considering valuation and replication in (largely) model-free environments by relying on advanced machine learning models (Buehler *et al.*, 2019; among others).

Second, I reconsider the foundations of the approach proposed here vis-à-vis the foundations of the “risk-neutral methodology” FP23 propose to “shift away from.” I echo some of the shortcomings of conventional valuation FP23 point out, agreeing that naïve application in the context of long-term liabilities can lead to problematic implications. However, I argue that the required shift runs deeper than what the authors argue for. Incorporating aspects like a “liquidity premium” or multiple values for seemingly identical payoffs – as the authors claim their approach

does – requires rethinking some of the foundations of asset pricing in market equilibrium, which is the pursuit of a growing literature on the macroeconomic relevance of financial frictions (Gromb & Vayanos, 2010; Brunnermeier *et al.*, 2013; among many others).

## 2. Benchmark 1: The Financial Production Model

As FP23 make clear, in conventional models including Black-Scholes and other standard, frictionless, complete-market derivative-pricing models, their approach yields the same value as risk-neutral valuation. Hence, it is important to emphasise that the strikingly lower values for financial and annuity products displayed in their figures and tables are cast within a rather specific model, namely “a stylised version of the minimal market model” provided in Section 5.3 (equations (13)–(17)). Whether this is a suitable and superior model for describing the S&P500 total return index is difficult to say and arguably would require a debate on financial-econometric grounds – although the guidance for pension funds or insurance companies to invest all their funds into the S&P500 and not consider the many other assets they have access to seems questionable. Yet, the authors do not seem to argue for a general shift towards that particular model but rather they argue for a fundamental shift in methodology towards the cheapest way to replicate given financial payoffs giving up the chains of specific stochastic models that allow for risk-neutral measures.

This objective relates to a recent literature on “deep hedging” that considers the cheapest way to replicate financial claims using a (largely model-free) machine learning framework (Buehler *et al.*, 2019). The ingredients to the deep hedging approach are as follows: A market scenario generator, a loss function, market frictions – which can include trading costs, risk limits, etc. – and trading instruments – which can go beyond simple primary assets but can also include liquid options traded in financial exchanges. Given a payoff and using generated market scenarios as the data, the approach trains a deep neural network machine learning model to obtain a minimal indifference price and the optimal trading strategy resulting in minimal loss, which in the context of Buehler *et al.* (2019) are cast via a convex risk measure. In line with FP23, Buehler *et al.* (2019)’s deep hedging approach “need no [...] equivalent martingale measure model,” and focuses “modeling effort on realistic market dynamics” and the replication/“hedging performance” (Buehler *et al.*, 2019). Hence, the deep hedging approach might be interpreted as a machine learning-based version of the production approach FP23 advertise.

There is a catch, however. In Buehler *et al.* (2019), the market scenarios are simulated by conventional financial models, which the authors later sought to extend to statistical market models, in line with FP23’s focus on the *real-world probability measure*. Yet, a naïve application of deep hedging using S&P500 data to almost any derivative results in the same “hedging” strategy: going long the index and selling puts (Buehler, 2022). As Buehler *et al.* (2021) explain: “Under this measure, we will usually find statistical arbitrage in the sense that an empty initial portfolio has positive value. This reflects the realities of historic data: at the time of writing the S&P500 had moved upwards over the last ten years, giving a machine the impression that selling puts and being long the market is a winning strategy. However, naively exploiting this observation risks falling foul of the ‘estimation error’ of the mean returns of our hedging instruments.”

The connection to FP23 is that their approach is subject to the same potential fallacy as the naïve machine learning algorithm. Starting with nothing and being able to generate a positive terminal value is problematic for a derivative valuation approach. In the setting of FP23’s Sections 5.1–5.4, if agents borrow at the treasury rate and invest the borrowed amount in FP23’s fair zero-coupon bond, this empty initial portfolio results in a positive value. Indeed, FP23 argue that having non-fair securities, that is, securities that trade below their minimal replication value, is crucial for their considerations.

Pointing to the argument that this represents financial market “reality” as shown in FP23’s figures is problematic since these figures are in the context of a single path of data and since “the

drift is a very difficult thing to estimate” – “if there is something in the data that you haven’t seen, there is a problem” (Buehler, 2022). The model producing the impression of being able to systematically beat long-term securities such as treasury bonds is a concern to the protagonists of the deep hedging approach at least, and their more recent work in Buehler *et al.* (2021) addresses this concern “by removing the drift.” This refined approach results in “clean” hedging strategies that are not “polluted by the trading strategy trying to make money from statistical arbitrage opportunities.” This presents a chasm to the approach in FP23.

A response to arbitrage arguments like the strategy outlined above (short treasury bond and long fair zero-coupon bond) is provided in FP23 via the basic Assumption 1. Since the (finite) benchmark portfolio is the best-performing portfolio, the argument goes, the existence of the strategy laid out above is ruled out by assumption. FP23 present this as an accomplishment of ruling out “economically meaningful arbitrage” in Section 3. But this strategy does seem viable in the context of Section 5, so that it could also be interpreted as a contradiction.

### 3. Benchmark 2: Macro-Finance Models

The figures and values provided in FP23 are striking. Furthermore, some of the implications of their analysis are appealing as they line up with various stylised facts related to financial/insurance markets that conventional models have difficulty explaining. Long-term investors often profess to exploiting liquidity premiums, which FP23 argue their approach incorporates (Section 5.2). Target date funds are popular for saving in retirement, whereas few individuals solely hold long-term bonds in their portfolios; the FP23 approach substantiates target date strategies (Section 5.1). More broadly, as the authors argue, equity returns are very high through the lens of long-term diversification, and, related to that, the risk-free rate is very low relative to equity returns over the long run (see FP23 Figure 2 and the corresponding discussion).

However, I do not agree with FP23 that the key issue is that conventional valuation assumes the existence of a risk-neutral measure. Indeed, I would argue that the existence of a risk-neutral pricing measure is not an *assumption*, as stated in FP23 several times, but a (mathematical) *consequence* of underlying assumptions, notably the absence of market frictions and the absence of arbitrage (Duffie, 2010). The question of how much given payoffs, such as the dividend stream produced by investment into a stock or a stock index, are worth generally is answered in the context of economic equilibrium. Farmers are endowed with the fruits of their lands, firms have technologies to convert labour and other input factors into goods, and consumers draw utility from their consumption of fruits and goods. Prices are the consequence of their trading activities that serve to optimise their own positions. And, under certain assumptions, the price system can be supported by a risk-neutral measure.

Within the context of these equilibrium asset pricing models, there are several puzzles that are closely related to some of the issues FP23 flag. Indeed, the so-called “equity premium puzzle” and “risk-free rate puzzle” are getting exactly at the huge disparity between long-run equity and treasury (risk-free) returns (Mehra & Prescott, 1985). As is detailed in this literature, the stylised facts are not inconsistent with standard financial theory per se, although it is difficult to reconcile them in Black-Scholes type economies with standard preferences. Approaches that can remedy the apparent inconsistency between equity returns and risk-free rates include the long-run risk approach by Bansal & Yaron (2004) or “extreme disaster” type models (Rietz, 1988; Barro, 2006), among others. For instance, the latter type of “extreme disaster” models proposes the possibility of extremal situations for the economy as a (theoretically consistent) way of explaining returns and risk-free rates, even in the context of FP23 Figures 1 through 5. Just because such an extreme economic disaster has never happened does not imply it is not going to happen in the future. So, it is not necessarily true that these stylised facts present conclusive evidence against the “classical risk-neutral methodology” with the fundamental assumption of absence of arbitrage per se, as FP23 argue in the last paragraph of Section 5.1.

However, there are examples of apparent arbitrage opportunities existing in certain markets even over longer periods, and these serve as the backdrop of a growing literature in financial economics regarding the “limits of arbitrage” considerations for explaining prices (Gromb & Vayanos, 2010). An important aspect of these models are *frictions* that affect financial intermediaries, which then reflect in equilibrium asset prices since intermediaries affect the allocation of resources and address liquidity mismatches. Thus, within these models, liquidity considerations and premiums become important (Brunnermeier *et al.*, 2013). As argued by Albrecher *et al.* (2018) and Bauer *et al.* (2023), capital market frictions are of key importance to insurance valuation.

FP23 suggest that their approach also features liquidity premia. However, unlike for the asset pricing models with financial frictions, the mechanism of how liquidity comes into play in their approach is opaque. Why is something more liquid and how does that manifest in the model, and why is it not arbitrated away?

To conclude, I concur with FP23’s perspective that empirical aspects of real-world markets, particularly with regard to long-term insurance liabilities, are at odds with some conventional approaches to valuation. Therefore, there is good reason to scrutinise some of the foundations underlying market-consistent valuation practices in the spirit of Black-Scholes and co. However, fundamentally valuation is not about super-martingales, probability measures, etc., but it is about agents facing risks and trading to offset these risks. From that vantage point, in my humble opinion, the approach presented in FP23 does not fundamentally fix or shift the theory underpinning valuation relevant to insurance liabilities, as the paper seems to suggest.

## References

- Albrecher, H., Bauer, D., Embrechts, P., Filipović, D., Koch-Medina, P., Korn, R., Loisel, S., Pelsser, A., Schiller, F., Schmeiser, H. & Wagner, J. (2018). Asset-liability management for long-term insurance business. *European Actuarial Journal*, 8, 9–25.
- Bansal, R. & Yaron, A. (2004). Risks for the long run: a potential resolution of asset pricing puzzles. *Journal of Finance*, 59, 1481–1509.
- Barro, R.J. (2006). Rare disasters and asset markets in the twentieth century. *Quarterly Journal of Economics*, 121, 823–866.
- Bauer, D., Phillips, R.D. & Zanjani, G.H. (2023). Pricing insurance risk: reconciling theory and practice. In *Handbook of Insurance*, 3rd Edition.
- Brunnermeier, M.K., Eisenbach, T.M. & Sannikov, Y. (2013). Macroeconomics with financial frictions: a survey. In D. Acemoglu, M. Arellano & E. Dekel (Eds.), *Advances in Economics and Econometrics, Volume II* (pp. 3–96). Cambridge University Press, New York.
- Buehler, H. (2022). Learning to trade - data driven quantitative finance. In Plenary Presentation at the 11th Bachelier Finance Society World Congress.
- Buehler, H., Gonon, L., Teichmann, J. & Wood, B. (2019). Deep hedging. *Quantitative Finance*, 19, 1271–1291.
- Buehler, H., Murray, P., Pakkanen, M.S. & Wood, B. (2021). Deep Hedging: Learning to Remove the Drift under Trading Frictions with Minimal Equivalent Near-Martingale Measures. arXiv preprint arXiv:2111.07844.
- Duffie, D. (2010). *Dynamic Asset Pricing Theory*. Princeton University Press, Princeton, NJ.
- Fergusson, K. & Platen, E. (2023). Less-expensive long-term annuities linked to mortality, cash and equity. *Annals of Actuarial Science*.
- Gromb, D. & Vayanos, D. (2010). Limits of arbitrage. *Annual Review of Financial Economics*, 2, 251–275.
- Mehra, R. & Prescott, E.C. (1985). The equity premium: a puzzle. *Journal of Monetary Economics*, 15, 145–161.
- Platen, E. (2006). A benchmark approach to finance. *Mathematical Finance*, 16, 131–151.
- Platen, E. & Heath, D. (2006). *A Benchmark Approach to Quantitative Finance*. Springer Science & Business Media, Heidelberg.
- Rietz, T.A. (1988). The equity risk premium a solution. *Journal of Monetary Economics*, 22, 117–131.