Drivers of Performance: Insights from a Member Outcomes Perspective

Melbourne Money and Finance Conference

February 2023

Drivers of Performance

Insights from a Member Outcomes Perspective

Andrew Cunanan and Nicholas Garvin*

APRA Working Paper

February 2023

Abstract

This paper empirically analyses drivers of performance – specifically, operating efficiency and investment returns net of investment costs and taxes. Funds experience economies of scale in operating efficiency, which have been stronger for growth through mergers, present throughout the full size range of Australian funds, and typically realised 1-2 years after growth occurs. Net investment returns relative to benchmarks show persistence, consistent with APRA's view that investment governance is fundamental to sustained performance. Relative returns are also weakly positively related to fund size.

*The authors would like to thank several others for helpful feedback and input, including Richard Brazenor, Leesa Chang, Miriam Escano, Wayne Fogarty, Mairead Loughlin, James Ma, Greg Simpson, Geoff Stewart, and members of the RBA Economic Research seminar series. The analysis and discussion in this paper have been produced by staff within APRA and do not necessarily reflect views of APRA as an organisation.

Introduction

Unacceptable performance in the superannuation industry hurts Australians' quality of life in retirement. In a perfect system, competitive forces would largely eradicate underperformance, as members would make fully informed decisions regarding their retirement savings and exit poor performing products as soon as they were identified. In reality, however, competitive forces within the superannuation industry are limited, as many members have lower than ideal levels of financial literacy and are not fully engaged with their retirement savings (e.g. Super System Review, 2010). Even for those members that are financially literate and engaged, identifying and exiting poor performing products is difficult due to limitations on availability of time and information.

In practice, and as part of APRA's broader mandate, APRA looks to hold Registrable Superannuation Entity (RSE) licensees ('trustees') to account for the decisions they make in respect of the retirement savings of their members. Through the superannuation Heatmaps and strengthened prudential framework (*Superannuation Prudential Standard 515: Strategic Planning and Member Outcomes*), APRA utilises its constructively tough supervisory approach to ensure that trustees are focused on improving outcomes for their members. APRA continues to see tangible improvements in the outcomes being provided to members across the entire industry.

APRA's Heatmaps present publicly available statistics on performance that can be easily benchmarked across the industry. The release of APRA's Heatmaps has increased the level of scrutiny on the member outcomes that trustees deliver, not only by APRA, but by industry commentators and members. While not directly designed for members, APRA's Heatmaps have made it easier for members to identify underperformance across the industry and make more informed decisions.[1]

This paper explores drivers of the performance measures in the APRA Heatmap. The focus is on two core components: 1) administration fees; and 2) net investment returns (NIR), defined as investment returns net of investment fees and costs and investment taxes. The analysis of administration fees focuses on drivers of the operating expenses that administration fees cover. The analysis of NIR focuses on drivers of returns within the trustee's control, in particular investment governance, and examines relationships between NIR and fund characteristics such as size.

Our analysis shows that fund size is a clear driver of performance, primarily through operating expenses, as operating expense ratios and size are robustly negatively related. The analysis also shows: merger-generated scale efficiencies have been relatively strong; efficiency gains appear achievable at all fund sizes, although inefficiencies are most evident for funds with less than $1 billion in assets; and, when growth has been driven by mergers and member movements, efficiency improvements have been realised around two years after

---

[1] The Heatmaps are available at Superannuation heatmaps | APRA.

the growth. We also find that NIR relative to benchmarks are persistent over time, with pockets of over and underperformance across the industry. The observed persistence generally aligns with APRA's supervisory observations regarding the quality of the investment governance practices of funds and, to a lesser extent, how governance relates to fund size.

An important caveat is that this analysis does not cover all aspects of performance. Administration fees and NIR are fundamental to performance, but other important components are less straightforward to measure quantitatively, such as quality of insurance, suitability of product offerings for target members, and effectiveness of customer service. Strong performance in these areas may in some instances justify higher operating expenses and/or lower overall investment returns. A related caveat for investment returns is that portfolio risk is difficult to objectively measure, but an appropriate portfolio risk profile is crucial for delivering good member outcomes.

Drivers of Administration Fees

This section analyses the operating expenses that administration fees cover, and builds on the previous analysis of fund sustainability in APRA (2022). Given the mechanical relationship between administration fees and operating expenses, the focus shifts immediately to operating expenses.[2] The four main empirical results are:

- Fund size is a robust driver of operating efficiency.
- Merger-generated scale efficiencies have been stronger than efficiencies from other growth.
- The mid-sized and large funds have not exhausted all scale efficiencies, and the highest inefficiencies are in funds with less than $1 billion in assets.
- Scale efficiencies lag growth by around two years, most evidently for growth through mergers and other member movements.

Expenses are examined at the fund level. This acknowledges costs that are not necessarily product-specific, such as staff, information technology and marketing.

Expense ratios have declined as the industry has grown and consolidated

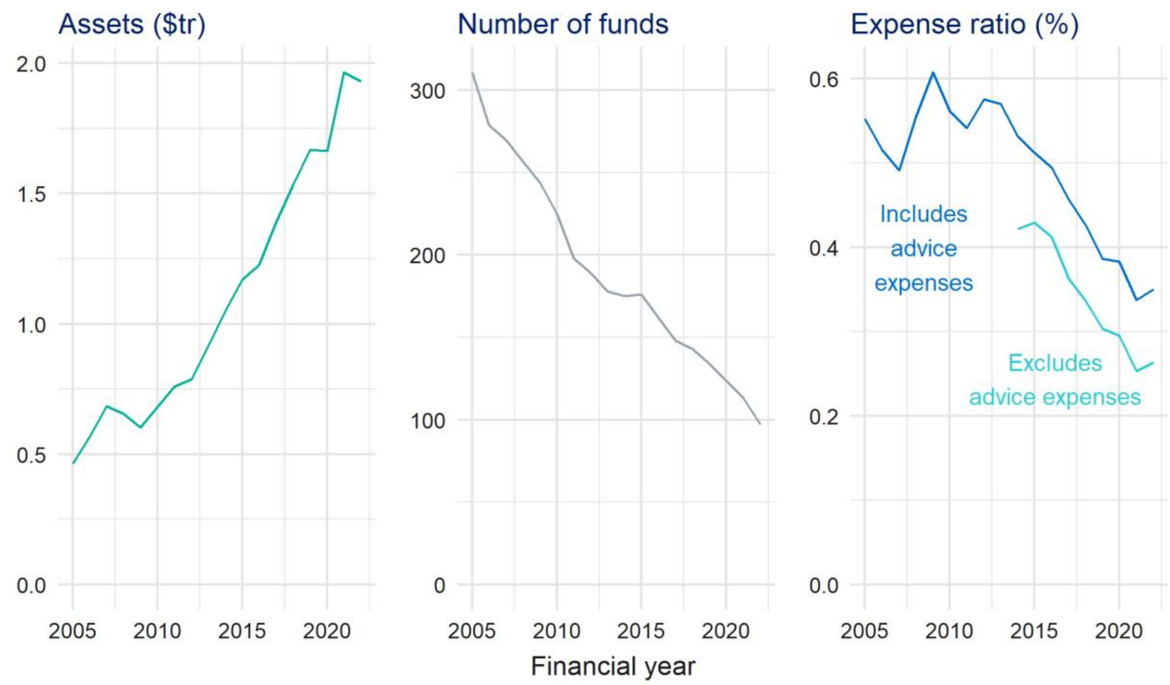
Over the last two decades, the superannuation industry has roughly quadrupled in size and the number of funds has dropped by around two thirds (Figure 1). Operational efficiency has also noticeably improved – in the last decade, the industry-wide expense ratio (annual operating expenses over assets) has almost halved.

---

[2] As expected, there is a robust and statistically significant positive relationship between representative fees for MySuper products and fund-level operating expense ratios (see Appendix).

The growth in the size of the industry and the reduction in the number of funds (fund consolidation) have clear drivers. Growth in the size of the industry has been driven by maturation of the superannuation system (e.g. Trinh et. al, 2019), rising nominal incomes, and the rising Superannuation Guarantee rate for mandatory employer contributions. The reduction in the number of funds has been facilitated, at least in part, by ongoing government reforms focused on increasing the quality of member outcomes being delivered by the industry.[3] One example was the recent introduction of the legislated Performance Test,[4] which was pre-empted by the Productivity Commission (2018a) when it recommended an elevated outcomes test for MySuper products to help achieve "much-needed consolidation in the super system".

Figure 1. Industry-wide growth, consolidation and efficiency



Notes: Excludes funds with <$50m assets for consistency with pre-2014 data.
Pre-2014 operating expenses data included advice expenses.
Sample of funds may differ from other available samples due to filters applied for data cleaning.

Previous work empirically ties the operating efficiency improvements evident in Figure 1 to industry growth and fund consolidation (Cummings, 2016; Productivity Commission, 2018b). Now that several years of detailed data are available, this section revisits the question and applies microeconometric tests to explore characteristics of the relationship.

---

[3] These include the 2009 Cooper Review (see Super System Review, 2010) and the 2018 Productivity Commission Inquiry (e.g. Productivity Commission, 2018a), among others.

[4] More information is available at The Annual Superannuation Performance Test | APRA.

Why would scale efficiencies exist?

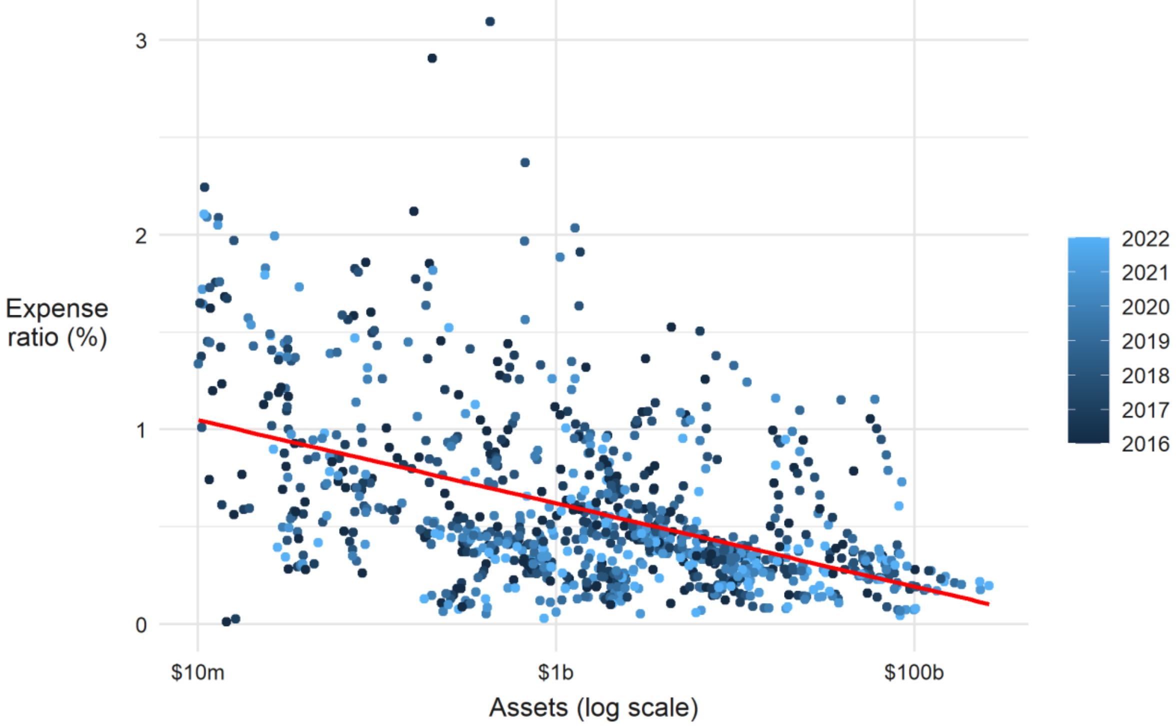
The data portray a clear negative relationship between expense ratios and fund size (Figure 2). There are several reasons why larger funds could operate at commensurately lower costs, including:

- Fixed costs in administration, such as setting up and maintaining IT systems. When administration is outsourced, these may be reflected in the provider's fee schedule.
- Fixed costs in staffing, or similarly, competitive advantages in attracting staff.
- Ability to insource business functions, such as administration, communications or investment management. This could result in business functions better tailored to the fund, and/or avoidance of mark-ups paid to service providers.
- Greater bargaining power over service providers, for fees and for service quality.
- Brand prominence, which can attract members and reinforce other scale efficiencies.

Nevertheless, some drivers of operating efficiency could be uncorrelated or negatively related to size. These could include business agility, technological sophistication, and control over governance processes and culture.

While larger funds are *likely* to be more efficient than smaller funds due to scale, it is important not to interpret the results to be a case of "small funds are always inefficient". This is not consistent with what APRA observes in practice, as a number of smaller funds are operating efficiently, also evident in Figure 2.

Figure 2. Expense ratios are negatively correlated with fund size



Notes: Red fitted line uses bivariate ordinary least squares (OLS). Entities under $10 million in assets are excluded.

How to estimate scale efficiencies in operating expenses

The following analysis of size and efficiency applies two econometric models, one static and one dynamic, to the panel of funds across financial years 2016 to 2022. Details of the models and the sample are in the Appendix.

The static model tests the existence, strength and functional form of scale efficiencies. It regresses the logged value of annual operating expenses on the logged value of assets, with a set of additional control variables that in some cases includes fund and/or year fixed effects. This log-log model is the accepted approach for estimating economies of scale (e.g. Greene, 2008). The coefficient on log assets represents the percentage change in operating expenses associated with a 1% change in size. A coefficient of less than one implies, for example, that an increase in size is associated with a declining ratio of expenses to size, meaning that economies of scale are present.

The dynamic model tests the *timing* of the effect of size on operating expenses. It regresses year-on-year percentage growth in operating expenses on contemporaneous and lagged year-on-year growth in assets. If, for example, the coefficient on the second lag of assets growth is 0.5, then 1% growth of a fund is associated with 0.5% growth in operating expenses in the second year after the growth. The sum of the contemporaneous and lagged assets-growth coefficients represents the cumulative effect on operating expenses of 1% growth in size, which is conceptually similar to the coefficient on size from the static model.

Estimated scale efficiencies

Across all variants of the static model, there are statistically significant economies of scale in operating expenses. The same is true if the size measure is changed to be the total number of accounts or the number of active accounts, instead of assets, but the assets measure of size fits the data best (see Appendix). For most models, the increase in expenses resulting from 1% growth is estimated at around 0.8% (Table 1), which is similar to previous cross-country estimates (e.g. Bikker, Steenbeck and Torracchi, 2017). As an example, this estimated effect indicates that if a fund were to grow from $20 billion to $30 billion, its expense ratio would decline by around 2 basis points (bps) (Figure 3).

Table 1. Coefficients on log assets from variants of the static model

	(1)	(2)	(3)	(4)	(5)
Standard fund fixed effects (FE)	0.82^^^	0.82^^^	0.50^^^	0.83^^^	0.50^^^
Fund FE with mergers separated			0.77^^^		0.75^^^
Control variables included	no	yes	yes	yes	yes
Fixed effects	none	none	fund	year	fund and year

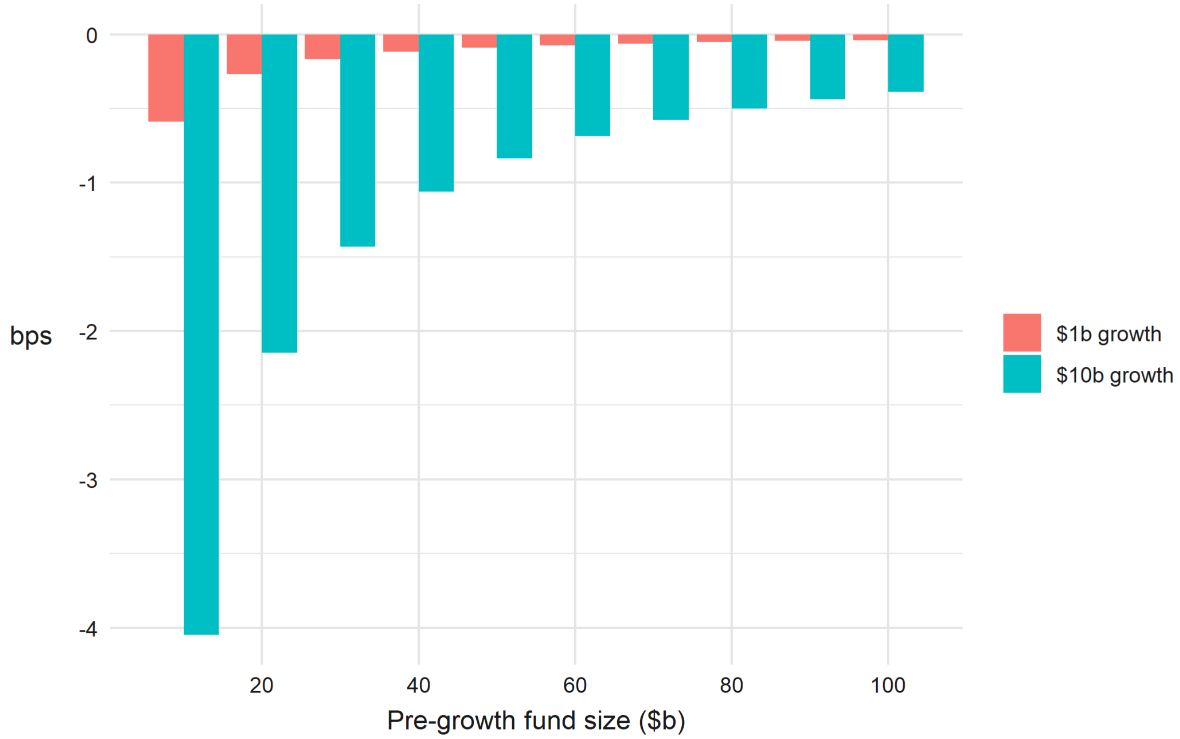
Notes: ^^^ denotes significantly different from 1 at 99%. Details of sample and specifications in Appendix.

Merger-generated growth has brought additional efficiency gains

To investigate the contribution of mergers to scale efficiencies, we modify the static model to remove size changes caused by mergers. This involves breaking up the fund fixed effects each side of a merger, which controls for any level shift in size and operating expenses that coincided with each of these mergers (see Appendix for details). When this modification is applied, estimated economies of scale are weaker – the coefficient moves from around 0.5 to

0.75 – but still present and statistically significant (Table 1, comparing the first and second rows).

Figure 3. Estimated efficiency gains from non-merger growth



Notes: Estimated declines in operating expenses based on coefficient of 0.8. Assumes expense ratios start at size-conditional sample averages.

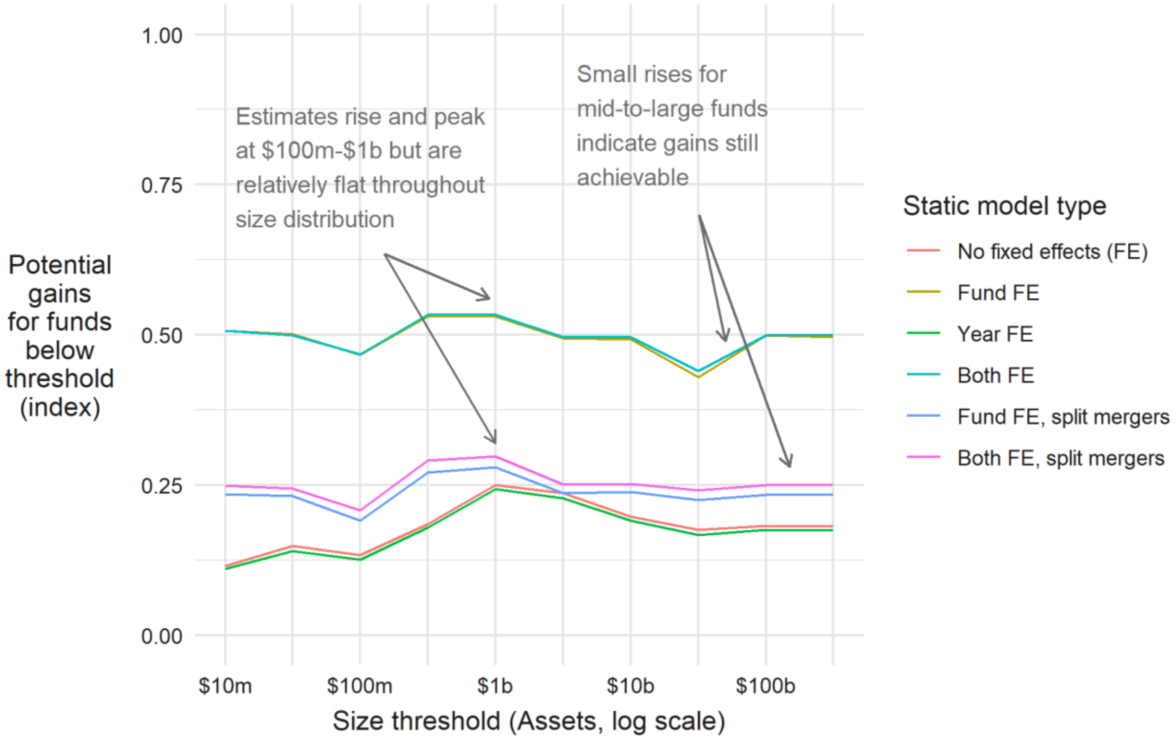
The weaker estimated economies of scale is evidence that scale efficiencies generated by mergers have been stronger than scale efficiencies driven by other sources of growth. This could be because mergers typically have the specific goal of improved performance, and merger partners may have been selected where the potential efficiencies and synergies have been strongest. This result represents mergers in the 2016-22 sample and should only be generalised to future mergers to the extent that they are similar in nature. Other APRA analysis is investigating methods for directly analysing merger effectiveness.

Potential efficiency gains are still present for the largest funds, but highest for funds with less than $1 billion in assets

To understand the sizes at which scale efficiencies have been strongest, we modify the static model to estimate the size coefficients for only funds below certain sizes. This involves interacting logged assets with an indicator variable for funds below a size threshold and allowing funds above the threshold to be captured by a separate coefficient (see Appendix for details).

Estimated scale efficiencies are still present at the largest fund sizes, and estimated potential gains are largest for funds with under $1 billion in assets (Figure 4). Overall, however, scale effects are relatively uniform across sizes – varying less than 0.15 in any given model – and statistically significant for all subsamples. Estimated efficiencies do not decline when funds above $32 billion and above $100 billion in assets are included, indicating that these funds are still achieving scale benefits. The peak estimated efficiencies are for funds below $1 billion, and estimates noticeably rise when entities above $100 million and $320 million are included, suggesting that the largest unrealised gains are in that region.

Figure 4. Economies of scale below various size thresholds



Notes: Efficiency gains index is one minus size coefficient from static model. All plotted indexes are statistically significant.

Bikker (2017) proposes generalisations to the static model that can detect potential diseconomies of scale at large fund sizes. Applying these generalisations on Australian data indicates an absence of diseconomies in operating efficiency in the current size range of Australian funds (see Appendix). This is consistent with the abovementioned finding that the largest Australian funds are still achieving scale benefits, and similar to the Bikker (2017) finding for Dutch pension funds.

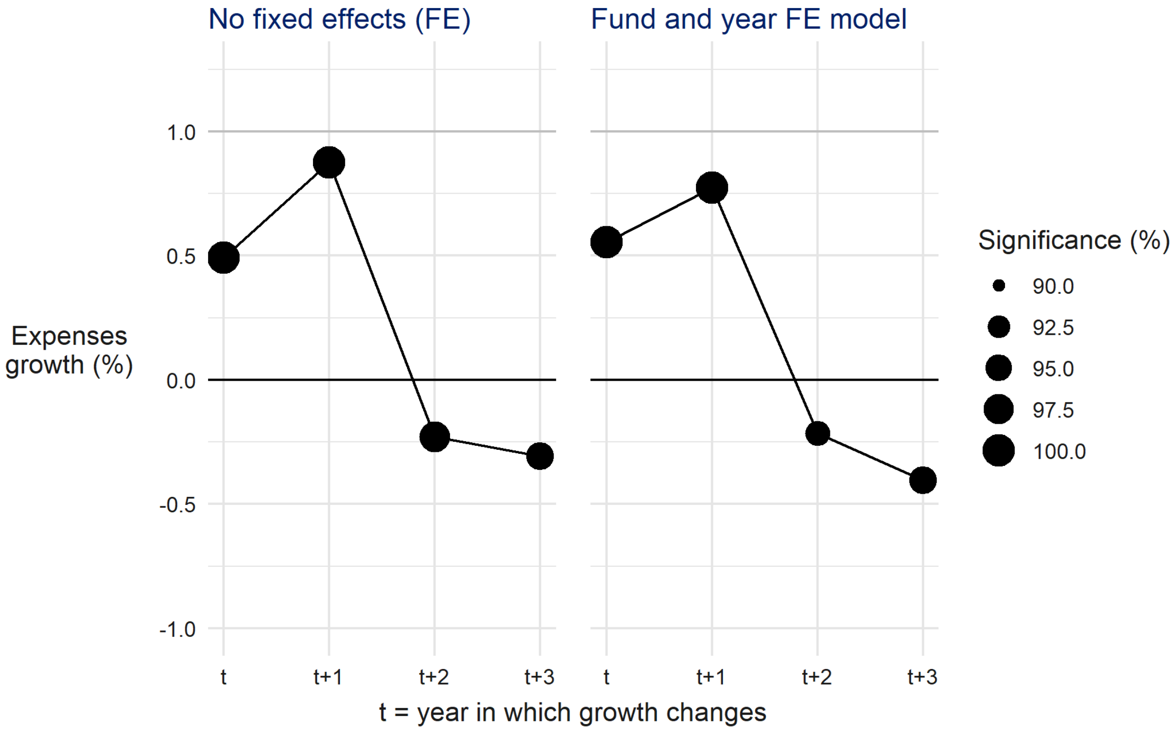
Efficiencies arrive on average two years after the growth

The dynamic model estimates that efficiency gains typically come two years after the year in which the fund growth occurs (Figure 5 reports two model variants with similar results). The largest rise in year-ended expenses occurs in year one, which is unsurprising because, for example, in some instances the fund growth would have occurred towards the end of year zero. After this peak, the dollar value of year-ended expenses declines in the following years. The estimated cumulative effects of 1% growth on expenses after three years are 0.83% and 0.71% for the two models, not far off the estimates from the static model. Modifications of the dynamic model indicate that this timing primarily reflects the effects of growth through mergers and through rollovers, and that natural growth through investment returns generates efficiencies faster (see Appendix). A caveat to these findings is that six periods (financial years 2017-22) is a relatively short time dimension for a dynamic model, so the results are best interpreted as outcomes in that specific period.

Other determinants of operating expenses

The control-variable coefficients in the static model demonstrate that size is not the only driver of expenses. For example, holding size constant, expenses are also increasing in the number of investment options offered and in the proportion of assets in defined benefit products. This highlights that higher expenses do not always imply lower operating efficiency; for example, defined benefit products may provide beneficial member outcomes that exceed the additional costs. APRA gauges these types of outcomes through its engagement with trustees. The Heatmap metrics on administration fees provide a focal point for conversations with trustees, from which nuanced discussions on member outcomes can progress.

Figure 5. Estimated effect of 1% RSE growth on expenses



Notes: Dots are statistical significance of difference from zero.

Drivers of net investment returns

Net investment returns (NIR) are largely driven by market movements, beyond the control of trustees. However, over the medium to long term, some funds repeatedly deliver higher returns than others. Differences in quality of investment governance likely explain at least some of this pattern, because investment governance frameworks cover key drivers of returns such as processes for selecting investments and investment managers, and for managing potential conflicts of interest. APRA's expectations for sound investment governance are outlined in *Superannuation Prudential Standard 530: Investment Governance* and related APRA guidance. The Heatmap investment metrics, some of which are analysed below, assist APRA in its supervision of investment governance. This section discusses the role of investment governance, and analyses characteristics of returns such as their persistence and relationship with fund size.

The importance of investment governance

While good investment governance is not necessarily a direct driver of returns in the short term, over time it supports ongoing investment performance by shaping the environment in which investment decisions are made. A sound investment governance framework (IGF) establishes, for example, clear accountability of investment outcomes, a culture of risk awareness, and robust processes for making timely investment-management decisions.

Whether ongoing investment performance is achieved depends on the Board's capability to establish and maintain a sound IGF. The Board's role includes setting the investment strategy, which covers processes around the selection of investments and investment managers, deciding how portfolio diversification will be achieved, and setting the processes for due diligence in investment decisions. A good IGF will also include robust performance-measurement and monitoring processes. This enables trustees to understand how their investments added or detracted value, to manage underperforming assets in a timely manner, and to review the strategy and its risks in response to changing market and economic conditions.

APRA has observed, through its supervision activities relating to the Heatmap and Performance Test, that trustees that it assesses as having weaker investment governance have also tended to deliver weaker returns over time, though there is substantial variation around this pattern and counterexamples exist.

Measuring returns

The following analysis focuses on NIR relative to strategic asset allocation (SAA) benchmarks for MySuper products (sometimes referred to as 'relative returns'). MySuper forms an important market segment, and the relative homogeneity of MySuper objectives facilitates comparison across products. The SAA benchmark – as included in the Heatmap and the legislated Performance Test – is used here as a proxy for what could be earnt from a hypothetical completely passive portfolio with the same risk profile. In other words, NIR relative to SAA benchmarks represent the value that the trustee generates through the implementation of its investment strategy. It is not a perfect proxy, in part because it does not account for portfolio risk within asset classes, and because setting the SAA is itself a decision of trustees. Nonetheless, arguably there is no perfect proxy, and the SAA benchmark has the advantages of having a clear interpretation and being transparently measured.

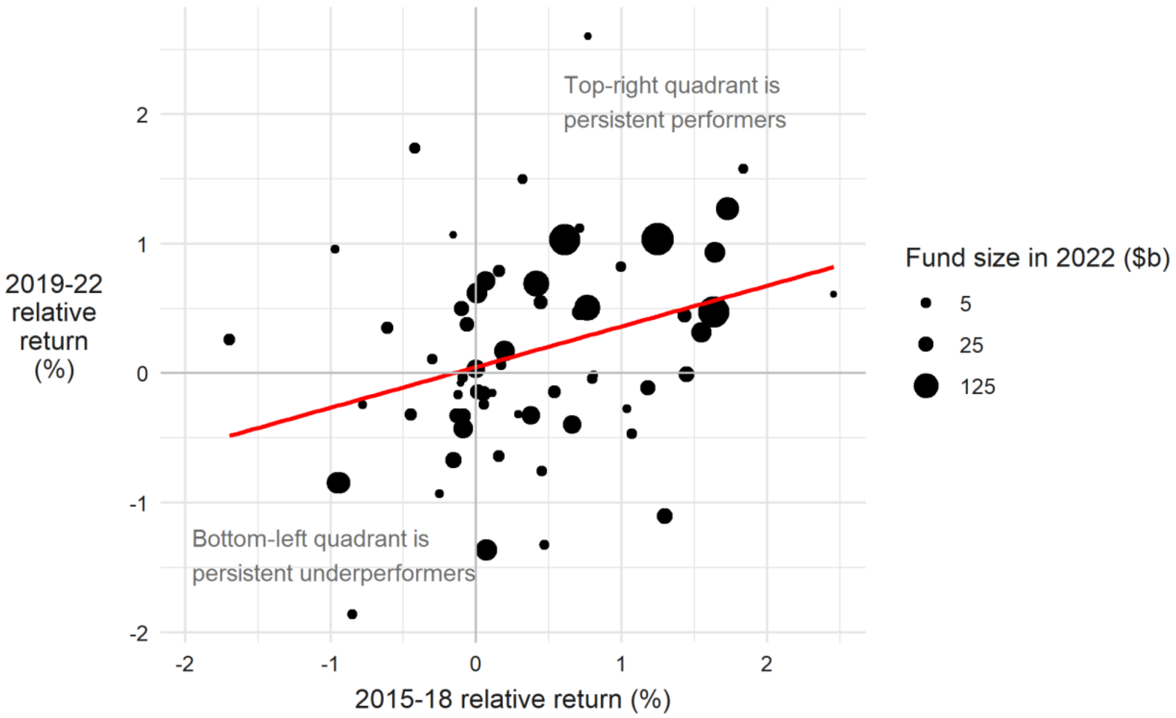
Over- and under-performance in relative returns is persistent

NIR relative to SAA benchmarks display visible persistence over the medium term (Figure 6), despite high variance over shorter-term horizons. Products with comparatively high relative returns across 2015-18 also had, on average, high relative returns across 2019-22. The positive relationship tends to hold when selecting different horizons within the 8-year sample, and is statistically significant at 95% confidence in a bivariate cross-sectional regression of 2019-22 relative returns on 2015-18 relative returns (see Appendix for details). Therefore, while each product has variation in relative returns across different time periods, there is an underlying level of performance that tends to persist across time, consistent with ongoing quality of investment governance.

Relative returns are weakly positively related to fund size

NIR relative to SAA benchmarks also have a slight positive relationship with fund size (Figure 7). This is broadly consistent with previous results in Australia (Cummings, 2016); results overseas are mixed (e.g. Amihud and Goyenko, 2013, and Pástor, Stambaugh and Taylor, 2015). Regressions of relative returns on logged value of assets produce p-values that vary between 0.01 and 0.13 depending on what other controls are included and whether the data are quarterly or annual (for details see Appendix). These estimates suggest that a doubling of fund size is associated with an annual improvement in relative returns of around 5 basis points. This relationship could be driven by scale efficiencies, such as fixed costs in internal investment management or greater negotiating power for outsourced investment management services. However, larger funds are also likely to have additional resources to achieve stronger investment performance through quality of investment governance, such as higher salaries to attract board members and investment staff, and access to specialist advisors such as tax specialists.

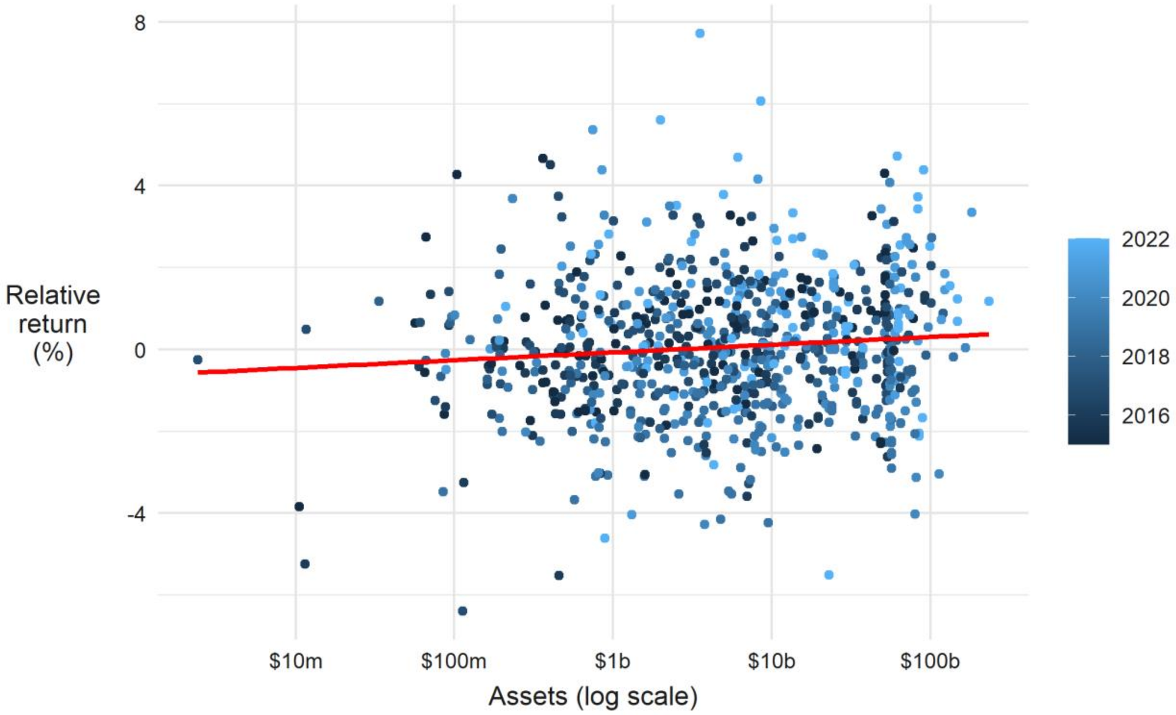
Figure 6. Returns relative to benchmarks are persistent



Notes: Red fitted line uses bivariate OLS.
Sample of all MySuper products in the market as at 30 June 2022.
Returns are relative to SAA benchmark and annualised.

Another reason why larger funds may be able to achieve better returns is a greater ability to access non-standard unlisted investments. Previous studies of superannuation find that unlisted or illiquid assets bring higher returns that more than compensate for the higher investment expenses incurred (Chant West, 2014; Cummings and Ellis, 2015). Direct investment in unlisted assets involves a high level of fixed costs and internal resourcing that, for smaller funds, may not be justified by the expected returns. These costs relate to deal sourcing, due diligence, legal documentation, capital structuring and shareholder engagement, and are typically lower or unnecessary when investing in listed assets. Unlisted assets can be invested in indirectly, though this involves an additional layer of external manager costs and still requires a reasonable level of internal resourcing to conduct manager due diligence. Passive investment is often not an option, unlike for most listed asset classes where benchmark returns can be achieved at relatively low cost, even for smaller allocations.

Figure 7. Annual returns relative to benchmarks and fund size



Notes: Red line is OLS fitted line.
Full sample of MySuper products since inception in FY 2015.
Returns are relative to SAA benchmark.

Conclusion

This paper analyses drivers of performance as measured by the APRA Heatmap. The Heatmap reports measures related to administration fees, investment returns and investment fees, to lift transparency and assist APRA in holding superannuation trustees to account for the member outcomes they deliver. The focus in this paper is on drivers of the operating expenses that administration fees cover, and of investment returns net of investment fees and costs and investment taxes. These drivers are analysed through visualisations and microeconometric analysis, and are discussed in the context of APRA experience.

The empirical analysis shows that fund size is a clear driver of performance, primarily through operating expenses. Efficiencies in operating expenses have been achieved through organic growth and through mergers, although in our sample the merger-generated efficiencies have been on average larger. Efficiencies through growth appear achievable at all fund sizes across the current industry size range, and largest for funds below $1 billion in assets. When funds grow through mergers and other member movements, the operating efficiencies tend to be realised a year or two later.

We find that differences in net investment returns relative to benchmarks across MySuper products have tended to persist over time, with pockets of over and underperformance across the industry. The observed persistence aligns with APRA's supervisory observation that the quality of investment governance is a crucial driver of returns. Fund size is also a determinant of relative returns. Size potentially allows additional resources to be allocated to investment governance, and can bring other efficiencies such as the ability to more effectively invest in unlisted asset classes.

References

Amihud, Y., & Goyenko, R. (2013). Mutual fund's R^2 as predictor of performance. *The Review of Financial Studies*, *26*(3), 667-694.

APRA (2022). 2021 Heatmaps – Sustainability of member outcomes. *APRA Technical Paper, March 2022*.

Bikker, J. A. (2017). Is There an Optimal Pension Fund Size? A Scale-Economy Analysis of Administrative Costs. *Journal of risk and insurance*, *84*(2), 739-769.

Bikker, J. A., Steenbeek, O. W., & Torracchi, F. (2017). The impact of scale, complexity and service quality on the administrative costs of pension funds: A cross-country comparison. In *Pension Fund Economics and Finance* (pp. 41-76). Routledge.

Chant West (2014). A low fee diet is no guarantee of healthy returns. https://www.chantwest.com.au/resources/a-low-fee-diet-is-no-guarantee-of-healthy-returns/

Cummings, J. R. (2016). Effect of fund size on the performance of Australian superannuation funds. *Accounting & Finance*, *56*(3), 695-725.

Cummings, J. R., & Ellis, K. (2015). Risk and Return of Illiquid Investments: A Trade-off for Superannuation Funds Offering Transferable Accounts. *Economic Record*, *91*(295), 463-476.

Greene, W. H. (2008). The econometric approach to efficiency analysis. In *The measurement of productive efficiency and productivity growth* (pp 92-250). Oxford Academic.

Pástor, L., Stambaugh, R. F., & Taylor, L. A. (2015). Scale and skill in active management. *Journal of Financial Economics*, *116*(1), 23-45.

Productivity Commission (2018a). Superannuation: Assessing efficiency and competitiveness. *Productivity Commission Inquiry Report, No.91, Overview.*

Productivity Commission (2018b). Superannuation: Assessing efficiency and competitiveness. *Productivity Commission Inquiry Report, No. 91, Technical Supplement 8: Economies of Scale in Superannuation.*

Super System Review (2010). Final report, part one: Overview and recommendations. Report from the Review into the Governance, Efficiency, Structure and Operation of Australia's Superannuation System, commissioned by the Australian Government.

Trinh K., Bott M., Stone J., Lavers E. and Luo N. (2019). Superannuation balances at retirement. *Australian Government Treasury Information Note.*

Appendix

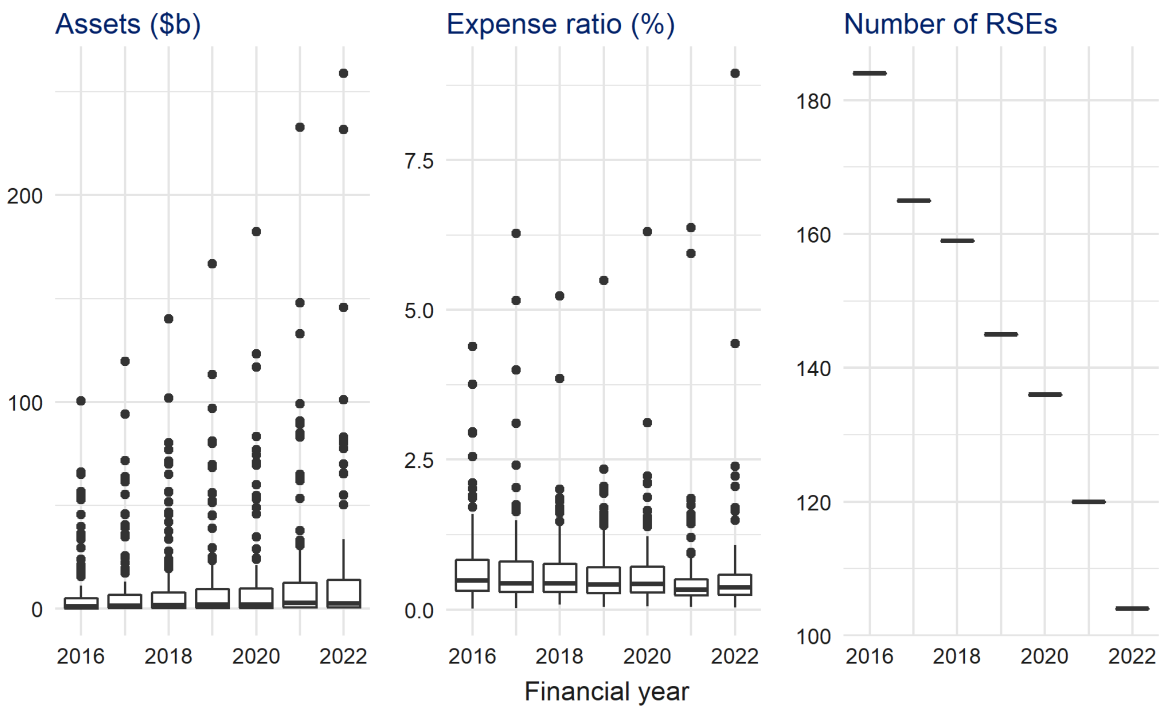
Drivers of administration fees – data and models

The analysis uses regulatory data collected by APRA, defined in the Superannuation Reporting Standards. 'Assets' is measured as item 21 *Net assets available for members' benefits* in SRF 320.0. Operating expenses are measured as item 10 *Administration and Operating Expenses* in SRF 330.0. The regression samples aggregate quarterly variables to the financial-year level, because operating expenses are highly seasonal within the year.

The definition of 'fund' excludes exempt public sector entities, pooled superannuation trusts and eligible rollover funds. The regression samples apply some filters to remove or truncate large outliers that appear to result from data reporting errors, or, for variables that are a proportion of assets, from sudden changes in the assets denominator due to merger-related activity.

The key variables and sample sizes are summarised in Figure A1.

Figure A1. Summary statistics for regression samples



Notes: Boxes illustrate the first, second and third quartiles.
Dots are values > 1.5x the interquartile range from the top or bottom quartile.

Administration fees and operating expenses

To test the relationship between administration fees and operating expenses, we use data on MySuper products for annual administration fees disclosed for a representative member with a $50 000 account balance. For funds that offer multiple MySuper products, fees are aggregated to the fund level by taking a weighted average, weighted by the assets in each MySuper product. This fee data is then merged with the fund-level dataset discussed above.

Denoting funds by i and financial years by t, the regression specification is

$$AdminFee_{it} = \beta ExpenseRatio_{it} + X_{it}\Gamma + \epsilon_{it},$$

where $AdminFee_{it}$ is the fee variable discussed above (expressed as a proportion of the account balance), $ExpenseRatio_{it}$ is the fund-level annual proportion of operating expenses to assets, and X_{it} is a vector of control variables. For each of four combinations of fixed effects – none, fund, year, and fund and year – we run two regressions, one with no other controls, and one including several fund-level control variables: proportion of assets in MySuper products; proportion of assets in defined benefit products; total reserves as a proportion of assets; and change in total reserves as a proportion of assets. Standard errors are clustered at the fund level.

The β coefficient estimates are reported in Table A1.

Table A1. Relationship between administration fees and expense ratios

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Expense ratio coefficient	0.17	0.19	0.21	0.21	0.19	0.20	0.24	0.17
Statistical significance	**	**	**	**	**	*	**	*
Control variables	no	no	no	no	yes	yes	yes	yes
Fixed effects	none	fund	year	fund & year	none	fund	year	fund & year
Sample size	548	548	548	548	448	448	448	448

Notes: * and ** denote statistically significant at 90% and 95%.

Static model details

Denoting funds by i and financial years by t, the baseline version of the static model is:

$$\log(OpExp_{it}) = \beta \log(Assets_{it}) + X_{it}\Gamma + \epsilon_{it} \,. \quad (1)$$

where X_{it} is a vector of control variables. The β estimate provides a statistical test of whether economies of scale exist:

$$\beta \geq 1 \;\Rightarrow\; constant\ or\ decreasing\ returns\ to\ scale$$

$$\beta < 1 \;\Rightarrow\; scale\ efficiencies\ exist.$$

The error term ϵ_{it} is clustered by fund (i) to deal with autocorrelation.

The vector of control variables X_{it} contains:

- Fund fixed effects (where mentioned).
- Year fixed effects (where mentioned).

- Dummies for fund type (corporate, industry, public sector or retail).
- Dummy for whether the fund operates under a commercial trustee (proxied by whether there are four or more funds under the same trustee).
- Logged number of investment options offered.
- Advice expenses as a proportion of assets.
- Investment expenses as a proportion of assets.
- Investment returns net of investment expenses (as a proportion of assets).
- Benchmark returns calculated by applying the SAA benchmark methodology to the fund-level actual asset allocation.
- The proportion of investments that are held in MySuper products.
- The proportion of balances that are held in defined benefit products.
- Total reserves as a proportion of assets.
- Proportions of investments held in:
    - Cash or fixed income asset classes (proxying defensiveness of portfolio).
    - Foreign asset classes.
    - Unlisted asset classes.
- Proportions of member accounts in each of the 11 age brackets in item 5 of SRF 610.0.

The fit of the static model is best when assets is used as the size measure. Table A2 reports the R squared when replacing $Assets_{it}$ in equation (1) with the total number of accounts or the number of active accounts. Assets fit the data best for all model specifications.

Table A2. Comparing the R squared of variants of the static model

	(1)	(2)	(3)	(4)	(5)
Size measure					
Assets	0.898	0.917	0.985	0.918	0.986
Accounts	0.829	0.889	0.984	0.890	0.985
Active accounts	0.832	0.891	0.984	0.892	0.985
Control variables	no	yes	yes	yes	yes
Fixed effects	none	none	fund	year	fund and year

Static model modifications

The first model modification separates the fund fixed effects either side of mergers. Mergers are identified as when: 1) the absolute net quarterly flow of benefits from successor fund transfers (SFTs) exceeds 2% of assets; and 2) the net quarterly flow of benefits from SFTs divided by the quarterly change in fund size excluding growth due to returns is greater than half. The second condition, which is only satisfied if net SFTs and fund growth (excluding returns) have the same sign, helps to ensure that instances of strong organic growth are not attributed to relatively small SFTs.

Table A3 summarises the number of mergers during the sample. Most funds were involved in no mergers, 24 were involved in one merger, and several were involved in three or more. If a fund closed by merging into another fund, its data series ends at that point in time.

Table A3. Incidence of mergers in sample

Mergers	Number of funds
0	158
1	24
3	5
2	2
4	1

The results presented are similar when changing the definition of 'merger' by increasing the 2% SFT threshold to 5%.

The second model modification estimates a separate size-expenses relationship for funds below a given size threshold. Define the variable $Small_{it}$ as indicator for whether the fund's assets are below some threshold $\overline{Assets}$:

$$Small_{it} = \begin{cases} 1 \; if \; Assets_{it} \leq \overline{Assets} \\ 0 \; if \; Assets_{it} > \overline{Assets}. \end{cases}$$

The static model is then modified as

$$\log(OpExp_{it}) = \beta^* \log(Size_{it}) * Small_{it} + \beta_{other} \log(Size_{it}) * (1 - Small_{it}) + X_{it}\Gamma + \epsilon_{it}.$$

In this model, the coefficient β^* captures scale efficiencies only for funds below the size threshold $\overline{Assets}$.

Dynamic model details

The dynamic model specification is

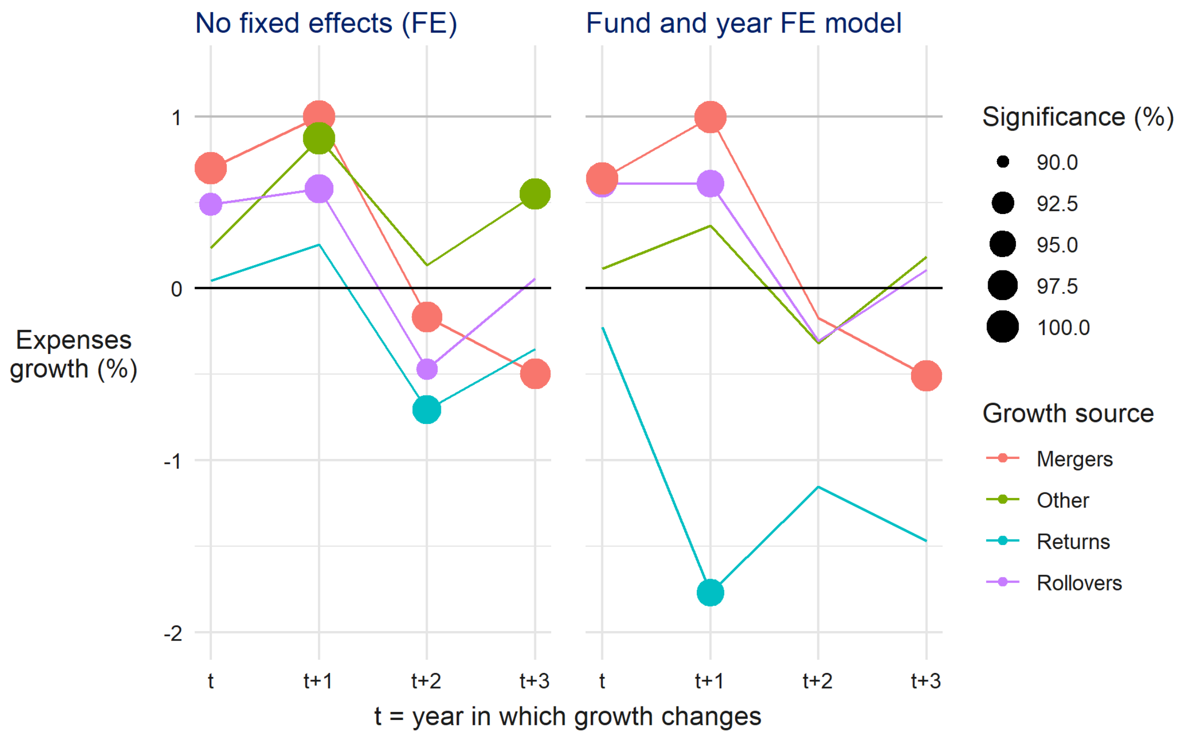
$$OpExpGrth_{it} = \alpha_i + \alpha_t + \sum_k \sum_{p=0}^{3} \beta_p^k * AssetsGrth_{it-p}^k + \epsilon_{it},$$

where operating expense growth ($OpExpGrth_{it}$) is the annual percentage growth in $OpExp_{it}$ and the cashflow variables ($AssetsGrth_{it}^k$) capture growth in assets. In the baseline dynamic model, k comprises a single growth type capturing total change in assets. In the modified dynamic model, k comprises four growth subcomponents that sum to total assets growth: investment returns; merger-driven asset changes; net inward and outward member rollovers; and all remaining growth, which is mainly contributions and lump-sum/pension payments. The α terms are fund and year fixed effects, which are excluded from some specifications.

The β_p^k coefficients estimate the reaction to growth of expenses in the immediate and following periods, and align neatly with the static model. For example, if $\beta_1^k = 1$, then 1% fund growth (of type k) is followed by a 1% increase in expenses the following year. More generally, $\sum_{p=0}^{3} \beta_p^k$ represents the cumulative percentage effect on expenses over three years. If most of the total effect occurred within three years, then this sum should be close to β from the static model.

Figure A2 reports results from the modified dynamic model.

Figure A2. Estimated effect of 1% entity growth on expenses



Dots are statistical significance of difference from zero.

In the static and dynamic models, the residuals (ϵ_{it}) are clustered by fund to deal with autocorrelation.

Identification of causality

In the static model, reverse causality is possible, because funds sometimes spend money to achieve growth, for example through improving websites and online interfaces to attract and retain members. However, it is not clear that this would bias β down from 1, which would depend heavily on the average cost efficiency of those growth strategies. Moreover, the dynamic model shows that efficiency gains lag growth, whereas reverse causality would produce the opposite outcome.

The Bikker (2017) model generalisation

Bikker (2017) suggests three types of generalisations to the static model, that can detect diseconomies of scale at large sizes. These models are referred to as the translog cost function, the unrestricted Laurent function (ULF), and the hyperbolically-adjusted Cobb Douglas cost function. For each of these, we estimate four variants: with and without the control-variable vector, and with and without fund fixed effects. For the ULF models without fund fixed effects, our estimation software (R with the 'plm', 'sandwich' and 'coeftest' packages) is not able to invert the matrices required for solving the estimates. Of the remaining 10 sets of estimates, the fitted cost elasticity is below 1 for all fund sizes from $10 million to $1 trillion in assets (assessing each power of 10 in that range). Therefore, there are no estimated diseconomies of scale in our sample.

Drivers of returns – data and models

The returns data cover MySuper products over the 2015-22 financial years (FY). The net investment return (NIR) and strategic asset allocation (SAA) benchmark data are the same data used for the annual performance test and the NIR relative to SAA benchmark of the 2022 APRA MySuper Heatmap.

Persistence regression specification

To test for persistence in NIR relative to SAA benchmarks ('relative returns'), relative returns for both the FY 2015-17 and 2018-22 periods are calculated as the annualised compounded four-year NIR minus the annualised compounded four-year SAA benchmark. Only products with the full eight-year history are included. Denoting MySuper products by i, and the two four year periods by 2017 and 2022, the ordinary least squares regression specification is

$$RelativeReturn_i^{2022} = \alpha + \beta RelativeReturn_i^{2017} + \epsilon_i.$$

Table A4 reports that FY 2018-22 relative returns are positively related to FY 2015-17 relative returns for the 65 MySuper products. The relationship is statistically significant at the 5% level (and close to being statistically significant at the 1% level).

Table A4. Persistence regression results

Intercept	0.000
	(0.640)
Slope	0.314
	(0.012)
R-squared	0.096
Sample size	65

Notes: p-values in parentheses

Size regression specifications

The regressions of relative returns on size are run on datasets of quarterly and annual frequency. Similar to the persistence regressions, annual relative returns are calculated as the compounded NIR minus the compounded SAA benchmarks.

The regression specifications are

$$RelativeReturn_{it} = \alpha_t + \beta \log\left(Assets_{it-1}\right) + X_{it}\Gamma + \epsilon_{it},$$

where α_t are time fixed effects and X_{it} is a vector of control variables. The industry average relative return has trended throughout the sample, so time fixed effects are important for ensuring that any correlation between this trend and the fact that funds have tended to grow is not spuriously caught by the coefficient of interest β. Product-level fixed effects are not included because the cross-product variation is the main variation of interest. The error term ϵ_{it} is clustered at the product level. The variable $\boldsymbol{Assets_{it}}$ is measured in dollars at the fund level rather than the product level, so, for example, if one fund offers multiple MySuper products, its assets mesaure is the same for each of those products.

Eight versions of this model are run, four at the quarterly frequency and four at the annual frequency. In each version, relative returns are measured in basis points and expressed on an annualised basis. The four model versions for a given frequency vary by the controls in X_{it}, with different combinations of: proportion of the portfolio invested in unlisted property/infrastructure; fund-type dummy variables (industry, retail, public sector and corporate); and investment fees (as a proportion of assets).

The coefficients at the quarterly frequency vary between 5.73 and 7.51, with p-values ranging between 0.05 and 0.13. The coefficients at the annual frequency vary between 8.01 and 10.08, with p-values ranging between 0.01 and 0.04.

The APRA Heatmap

APRA publishes a MySuper Heatmap annually to improve transparency in the industry and hold trustees accountable for outcomes they deliver to MySuper members.[5] It provides comparable information for various metrics across MySuper products including:

- the result of the annual performance test;
- investment return metrics over different horizons (net return, NIR, NIR relative to SAA benchmark, NIR relative to a simple reference portfolio);
- administration and total fees and costs metrics across different account balances;
- sustainability metrics including member accounts growth, net cashflow growth and rollover growth.

[5] See https://www.apra.gov.au/mysuper-product-heatmap-0.

The methodology used to produce these metrics is outlined in a methodology paper available on the APRA website.[6]

The NIR relative to SAA benchmark is used in this paper and aligns with the investment component used for the legislated annual performance test of MySuper products in Australia. Trustees report NIR and strategic asset allocation to APRA on a quarterly basis. The SAA benchmark is created by mapping the reported SAA to the covered asset classes of the annual performance test (beginning of quarter) and multiplying by the quarterly index returns for each asset class as adjusted for standard tax and fee assumptions (as specified in the legislation).

Another measure in the Heatmap, not used in this paper, is NIR relative to Simple Reference Portfolio. This metric seeks to measure the value-add achieved by trustees through strategic asset allocation decisions relative a passive portfolio with the same average exposure to growth versus defensive assets.

[6] See https://www.apra.gov.au/sites/default/files/2022-12/Methodology%20paper%20-%20MySuper%20Heatmap.pdf.