April 2022
TRP22-13

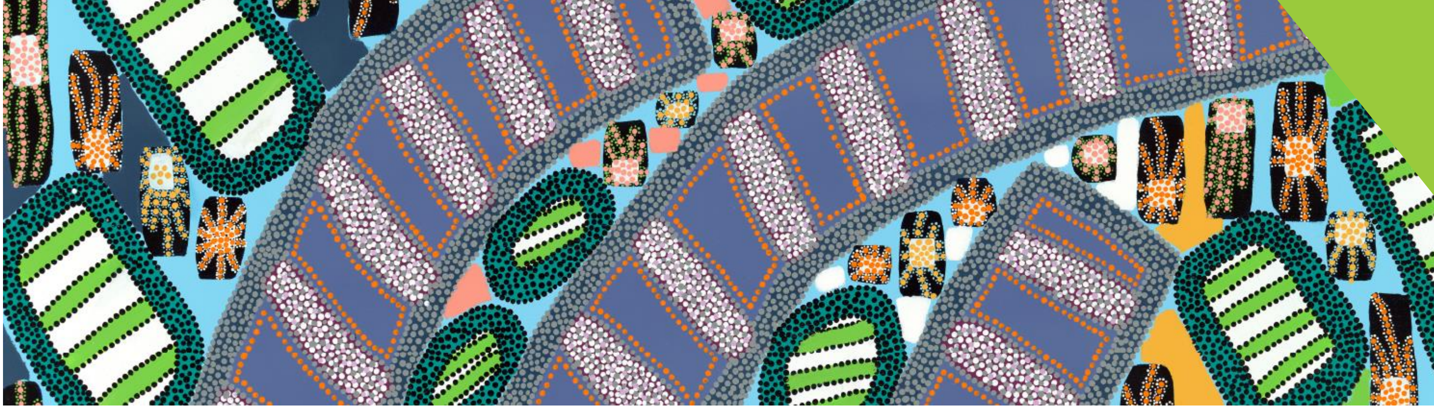
Holding periods of residential property buyers in NSW

Guilherme Bandeira[*], Michael Malakellis[**] and Michael Warlters[***]

[*] NSW Treasury

[**] KPMG Brisbane

[***] NSW Treasury



Acknowledgement of Country

NSW Treasury acknowledges that Aboriginal and Torres Strait Islander peoples are the First Peoples and Traditional Custodians of Australia, and the oldest continuing culture in human history.

We pay respect to Elders past and present and commit to respecting the lands we walk on, and the communities we walk with.

We celebrate the deep and enduring connection of Aboriginal and Torres Strait Islander peoples to Country and acknowledge their continuing custodianship of the land, seas, and sky.

We acknowledge the ongoing stewardship of Aboriginal and Torres Strait Islander peoples, and the important contribution they make to our communities and economies.

We reflect on the continuing impact of government policies and practices and recognise our responsibility to work together with and for Aboriginal and Torres Strait Islander peoples, families, and communities, towards improved economic, social, and cultural outcomes.

Artwork: ‘Regeneration’ by Josie Rose 2020

Contents

1. Introduction

The purpose of this paper is to provide information about how long purchasers hold residential property. The paper estimates probability distributions that indicate the likelihood that a property will be sold in each year of ownership. While half of all purchasers sell their property within 10 years of purchase, there is a long tail of people who hold their properties for substantially longer. The mean purchaser's holding period is about twice as long as the median. Our estimates include separate distributions for owner-occupiers and investors, as well as the combined distribution for all purchasers.

Section 2 estimates the distribution of holding periods of purchasers using transaction data obtained from Revenue NSW. The main technical challenge is one of censored data. The longest period for which we possess both the purchase and sale dates of individual properties is 17 years. Inevitably there are many properties for which it is not possible to determine the complete holding period. The lack of data on longer holding periods requires assumptions to be made about the shape of the tail of the distribution.

In Section 3, information derived from aggregate housing market data is used to specify mean holding periods. In Appendix 1, we show that the mean purchaser's holding period is the inverse of the share of the established dwelling stock sold each year. This information is used to provide an additional constraint on the tails of the holding period distributions estimated in Section 2.

In Section 4, we examine holding period estimates published by property data companies. CoreLogic and REA Group report estimates for the average holding period of properties sold in individual years. They are thus an annual measure, rather than the equilibrium or long-run measure that we have sought to estimate. While we do not know their precise methodologies, it appears their estimates of average holding periods are biased down because they do not address the challenge of censored data.

Section 5 examines survey data recording how long owner-occupiers have lived in their homes. Data on incomplete tenure (i.e., time between purchase and the survey date) can in principle be used to estimate a holding period distribution. For a variety of reasons, however, we conclude that the available surveys do not provide a reliable basis to estimate purchaser holding periods.

Section 6 presents our preferred distributions, showing the likelihood of sale in each year after purchase, for owner-occupiers, investors, and residential purchasers as a whole. We find the median holding period of NSW residential property buyers is 9.7 years and the mean is 18.8 years.

Theoretical background, methodological details and additional results are presented in Appendices.

2. Holding periods estimated with transaction data

We use administrative records of property transactions to measure the holding periods of purchasers (HPP). We obtained transaction data from Revenue NSW to conduct our analysis. The data set is based on transactions records created for the primary purpose of administering the state's property taxes, including transfer duties and land tax. Although comprehensive in many respects, the data have various limitations, which we discuss below.

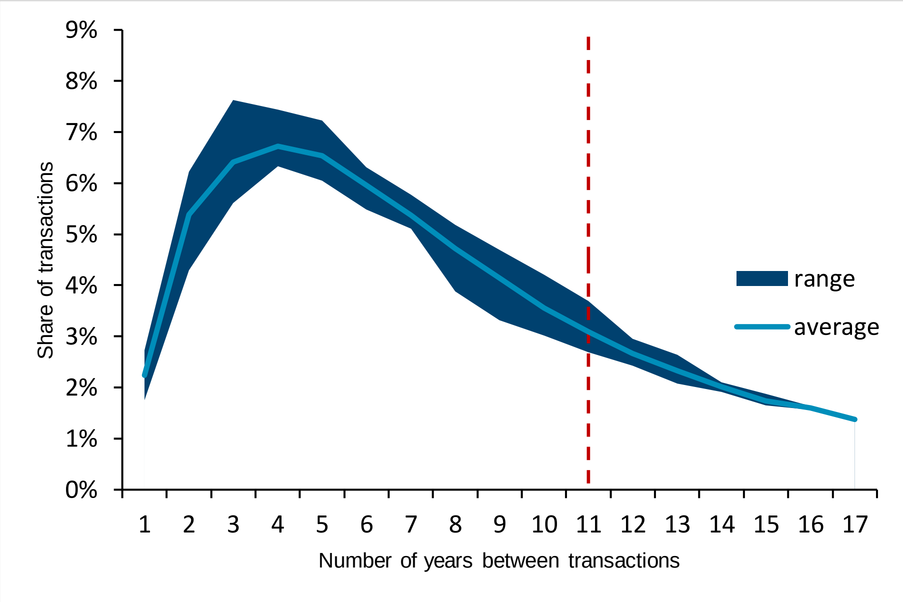
Revenue NSW transaction data can be used to identify properties that were purchased in a particular year and to measure the length of time between that point and when the property was next transacted. We examined transactions for residential properties purchased during the financial years 2004-05 to 2010-11. The data captures subsequent transactions of each of these properties through to 2020-21.

We refer to the holding period data associated with properties purchased in a particular financial year as a vintage. We measure the holding period of each property as the length of time from its purchase date (within the period 2004-05 to 2010-11) until the date it was next sold (during 2004-05 to 2020-21). The maximum holding period captured by the dataset is 17 years. A significant portion of properties purchased in 2004-05 (just over 29 per cent) had not been sold by the time our sample ends in 2020-21. For subsequent vintages the maximum holding period captured in the dataset gets progressively smaller, and a larger proportion of these properties were not resold within the period covered by the dataset. For example, just over 50 per cent of properties purchased in 2010-11 had not changed hands by 2020-21.

The holding period data for buyers of all types of residential properties is summarised in Table 2.1 and depicted in Figure 2.1. The maximum completed holding period for which we have observations *in every vintage* is 11 years. This is marked by the vertical dotted red line in Figure 2.1. This can be seen in Table 2.1 where the maximum completed holding period for properties purchased in 2010-11 is 11 years because we can only observe transactions conducted by 2020-21.

Except for the 2010-11 vintage, more than 50 per cent of the properties purchased in each vintage had been sold by 2020-21. For the 2010-11 vintage 49.6 per cent of the properties purchased in 2010-11 had been sold by 2020-21. Thus, in most cases the median HPP can be calculated without making assumptions about the unobserved part of the distribution. The median holding period for each vintage is reported in Table 2.2, along with an estimate for the 2010-11 vintage. The median HPP ranges from 8.8 to 11 years with an average of 9.7 years. In contrast to the median, it is not possible to estimate the mean HPP without making assumptions about the unobserved part of the distribution.

Figure 2.1: Distribution of holding periods for transacted NSW residential properties[(a)]



Note: (a) This distribution has been truncated at 17 years because this is the longest holding period identified in our dataset. That is, the earliest transactions we identify are in 2004-05 and the last completed year in our dataset is 2020-21.

Table 2.1: Holding period estimates - NSW residential transactions

Holding period	Vintages							
	2004-05	2005-06	2006-07	2007-08	2008-09	2009-10	2010-11	Average[(a)]
1	2.7%	2.3%	2.4%	1.9%	2.5%	2.1%	1.8%	2.2%
2	6.2%	6.2%	5.4%	5.7%	5.4%	4.7%	4.3%	5.4%
3	7.6%	6.1%	7.0%	6.6%	6.1%	5.6%	6.0%	6.4%
4	6.6%	7.4%	6.9%	6.5%	6.3%	6.4%	7.0%	6.7%
5	7.2%	6.6%	6.1%	6.1%	6.6%	6.7%	6.5%	6.5%
6	6.2%	5.8%	5.6%	6.2%	6.3%	6.2%	5.5%	6.0%
7	5.1%	5.1%	5.6%	5.8%	5.7%	5.2%	5.1%	5.4%
8	4.6%	5.0%	5.2%	5.2%	4.7%	4.5%	3.9%	4.7%
9	4.4%	4.7%	4.6%	4.3%	4.1%	3.7%	3.3%	4.1%
10	4.2%	4.0%	3.8%	3.6%	3.1%	3.0%	3.3%	3.5%
11	3.7%	3.4%	3.2%	2.9%	2.7%	3.0%	2.9%	3.1%
12	2.9%	3.0%	2.6%	2.4%	2.6%	2.5%		2.7%
13	2.6%	2.4%	2.2%	2.4%	2.1%			2.3%
14	2.0%	1.9%	2.1%	2.0%				2.0%
15	1.7%	1.9%	1.7%					1.7%
16	1.6%	1.6%						1.6%
17	1.4%							1.4%
H+[(b)]	29.1%	32.8%	35.7%	38.5%	41.7%	46.4%	50.5%	34.1%

Note: (a) The average is calculated by summing across all vintages the number of dwellings sold in a particular year of ownership and dividing this by the number of dwellings in all vintages. (b) These values capture observations with holding periods greater than the maximum holding period H observed in a vintage. For each successive vintage H gets smaller by one year. Thus, in 2004-05 H is 17 years meaning that for that vintage 29.1% of the observations are for hold periods of 18 years or more. For the 2010-11 vintage H is 11 years, meaning that 50.5% of the observations in that vintage are for holding periods of 12 years or more.

Table 2.2: Median holding period estimates

Holding period	Vintages							Average
	2004-05	2005-06	2006-07	2007-08	2008-09	2009-10	2010-11	
Median	8.8	9.2	9.3	9.5	9.7	10.6	11.2[(a)]	9.7

Note: (a) This is an estimate as only 49.5% of the properties purchased in 2010-11 had been sold by the end of the sample. We have assumed that the share of properties purchased in 2010-11 that will be held for 12 years is equal to the average of the comparable shares in the previous vintages (i.e., the shares of properties held for 12 years in each vintage).

The information in Table 2.1 indicates the shape of the HPP distribution up to 17 years. The remainder of the distribution remains unknown and will only be revealed gradually over time. However, we can estimate what the entire distribution might look like by fitting a theoretical distribution to the available data.

Fitting a theoretical distribution to the transaction data

We experiment with three alternative assumptions about the maximum holding period (65, 75 and 85 years). This range seems reasonable if we assume that holding periods are bounded by the natural life of owners and that ownership of a property by minors (under 18 years of age) is complex from a legal/tax standpoint. As an example, to hold a property for 85 years an adult purchaser would need to purchase at 18 and live beyond 100 years of age.

Bounded distributions that can accommodate the upper and lower limits are unlikely to adequately capture the shape of the empirical distribution for the holding periods that have been observed (i.e., the shape depicted in Figure 2.1). To deal with this issue we introduce flexibility into the estimation procedure by considering a mix of two theoretical distributions.

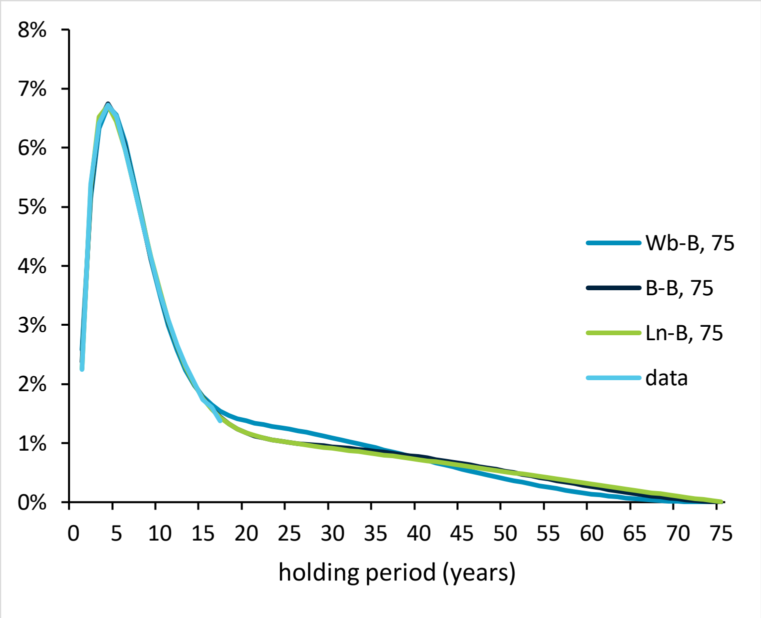
Details about our approach to fitting the mixed distributions to the data are set out in Appendix 2. To summarise, we specify two theoretical distributions and a weighting function and then choose the required parameters to minimise the root mean square error (RMSE) between the fitted distribution and the data. To maintain tractability, we consider only 2-parameter distributions, including Weibull, log-normal and beta distributions. The primary theoretical distribution, which is designed to target the left side of the HPP distribution, can be any of the three candidates considered. The secondary theoretical distribution targets the right side of the HPP distribution, which is assumed to be bounded. Only beta distributions are considered as candidates for the secondary distribution.

In Figure 2.2 we show fitted distributions for HPP against the part of the empirical distribution revealed by the transaction data (i.e., the average column in Table 2.1). We tested upper bounds of 65, 75 and 85 years. The fitted distributions of HPP shown in Figure 2.2 are based on a maximum holding period of 75 years, which was the upper bound providing the lowest RMSE. Additional results designed to show sensitivities to key assumptions are reported in Appendix 3.

Each fitted distribution is a mixture of two theoretical distributions with the mix identified by the labels. For example, "*LN-B,75*" refers to a mix where the primary distribution is log-normal, the secondary distribution is beta and the upper bound is 75 years. Analogously "*B-*

B,75” and “*W-B,75*” indicate mixes where the primary distribution is beta or Weibull and the secondary distribution is beta.

Figure 2.2: Fitted HPP distributions – All Property Purchasers



	Ln-B, 75	Wb-B, 75	B-B, 75
mean	17.2	16.0	17.0
median	9.7	9.6	9.7
RMSE	0.180%	0.524%	0.239%

The median HPP is similar across the various distributions reported and very close to the median of the empirical distribution. This reflects the fact that the estimated distributions fit the empirical distribution well over the part that contains the median.

The average HPP ranges from 16.0 years for the *W-B,75* distribution to 17.2 years for the *LN-B,75* distribution. The lowest RMSE is obtained for the *LN-B,75* distribution. If the average property transacts every 17.2 years, the probability of the average property transacting in any year is about 5.81 per cent. In Section 3 we consider this probability in the context of alternative estimates based on aggregate data.

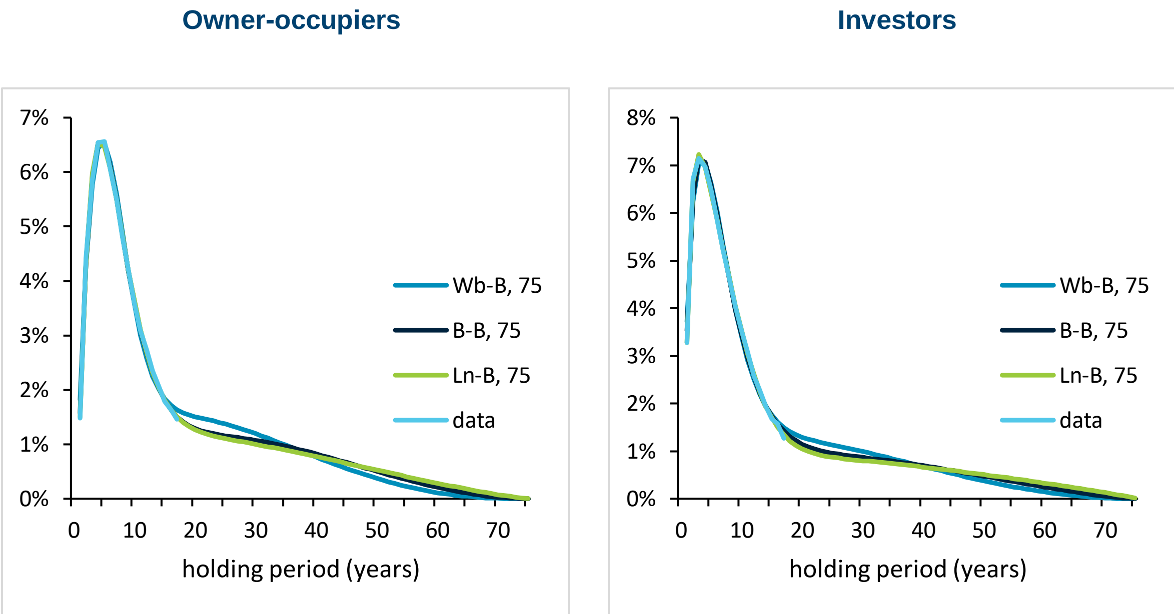
Holding periods of investors and owner-occupiers

The above analysis was based on transaction data for all residential properties. For land tax compliance purposes, Revenue NSW identifies whether dwellings are owner-occupied or investor-held. On the basis of these data flags, we find that across the 7 vintages of data the proportion of transactions accounted for by investors ranges from 35.8 per cent to 44.5 per cent with an average of 42.6 per cent. The Revenue NSW data flags may not capture all rental properties that are below the land tax threshold and not part of a larger property portfolio, however the resulting proportion is similar to that derived from home finance data, in which over the long run investors have averaged about 45 per cent of the number of purchasers.

We have repeated the distribution fitting exercise described above using the sub-samples of transaction data relating to owner occupiers and investors. In the left panel of Figure 2.3 we report the fitted distribution of HPP for owner occupiers based on different mixes of theoretical distributions. Analogous distributions for investors are presented in the right panel of Figure 2.3. The results reported in Figure 2.3 are comparable to the results for all property owners reported in Figure 2.2.

Focusing on the "*LN-B*" mixes we find that the mean and median HPP for owner occupiers are both about half a year longer than for all types of owners. This implies that the mean and median HPP for investors must be lower than for all types of owners, which is confirmed by the results reported in the right panel of Figure 2.3. The median HPP for investors is 1.4 years less than for owner occupiers, while investors' mean HPP is 0.9 years less than the mean for owner occupiers.

Figure 2.3: Fitted distributions differentiated by purchaser type



	Ln-B, 75	Wb-B, 75	B-B, 75
mean	17.6	16.3	17.2
median	10.3	10.2	10.3
RMSE	0.196%	0.515%	0.186%

	Ln-B, 75	Wb-B, 75	B-B, 75
mean	16.7	15.2	16.0
median	8.9	8.9	8.9
RMSE	0.171%	0.788%	0.537%

Summary

In this section we have used a mixture of theoretical distributions to fit administrative transaction data. The distributions that provide the best fit (i.e., lowest RMSE) for all buyers, owner occupiers or investors are derived from mixtures of log-normal and beta distributions, with an upper bound of 75 years. Under these specifications the average and median HPP are set out in Table 2.3:

Table 2.3: Estimated mean and median holding periods using transaction data

	Owner-occupiers	Investors	All purchasers
Mean	17.6	16.7	17.2
Median	10.3	8.9	9.7

Each distribution of HPP that we have estimated is a good fit for the observed part of the empirical distribution. However, the mass of the fitted distribution beyond holding periods of 17 years is determined by our assumptions relating to the maximum holding period, the candidate theoretical distributions considered and the approach to mixing these distributions. This is an important caveat on the reliability of estimates of key moments of the fitted distributions, including the average HPP. The shape of the right tail of the distribution will affect the estimate of the average HPP. Thus, the arbitrariness in the assumptions that determine the shape of the HPP distribution beyond 17 year holding periods will be reflected in the estimate of the average HPP.

3. Incorporating aggregate housing market data

The holding period distributions estimated in section 2 are based on Revenue NSW transaction data for holding periods of up to 17 years, but the tails of the distributions are determined purely by assumptions. With more than half of all property purchasers selling their dwellings within 10 years, the transaction data provide strong evidence about median holding periods. Estimates of average holding periods, however, could vary greatly depending on the assumptions used regarding the shape of the distribution's tail for longer holding periods.

To improve our estimate of the distributions' tails, we can use aggregate housing market information to identify mean holding periods.

Warlters (2022) has set out a home ownership model (reproduced in Appendix 4) in which, in equilibrium,

$$\beta_I = \frac{\alpha\tau}{\gamma} - \nu \qquad (3.1)$$

$$\beta_O = \frac{(1-\alpha)\tau}{(1-\gamma)} - \nu \qquad (3.2)$$

where:

β_I = share of investor-owned dwellings that are sold each year

β_O = share of owner-occupied dwellings that are sold each year

α = share of dwellings that are for sale each year that are purchased by investors

τ = annual transactions (including sales of established and new properties) as a share of the stock at the beginning of the year

γ = investor-owned dwellings as a share of the stock at the beginning of the year

ν = growth rate of the dwelling stock

Under the assumption of a stable purchaser holding period distribution, we show in Appendix 1 that the expected holding period is the inverse of the share of the stock sold each year. This finding must be adjusted for a growing stock, by focusing on the share of the established stock sold each year. That is,

$$\mu = \frac{1}{\tau - \nu} \qquad (3.3)$$

The same inverse relationship is observed in relation to the stock held by owner-occupiers and the stock held by investors, so that the mean holding periods of owner-occupiers and investors are the inverse of β_O and β_I respectively:

$$\mu_O = \frac{1}{\beta_O}$$

$$\mu_I = \frac{1}{\beta_I}$$

Table 3.1 sets out the assumed values for α, γ, ν and τ. Using these values and equations 3.1, 3.2 and 3.3, we find $\beta_O = 4.41\%$ and $\beta_I = 7.29\%$, which yields the following average holding periods for all purchasers, owner-occupiers, and investors:

$\mu = 18.7$ years

$\mu_O = 22.7$ years

$\mu_I = 13.7$ years

Table 3.1: Parameters used in the Home Ownership Model

Parameter		Value	Data source
Ratio of total transactions/dwelling stock	τ	6.54%	Average of Revenue NSW residential transfers 2000-01 to 2019-20/ Office of Local Government number of properties that pay residential rates.
Growth of dwelling stock	ν	1.19%	Assumed equal to average annual NSW population growth since 2000/01, using ABS data
Investor share of private dwelling stock	γ	32.9%	2016 Census data, excluding 'tenure-type not stated' and with adjustments for unoccupied private dwellings.[1]
Investor share of transactions	α	42.6%	Revenue NSW residential transactions data from 2004-05 to 2020-21; average.

These estimates of average holding periods are used to condition the HPP distributions estimated using the transaction level data in Section 2, providing an additional constraint on the assumptions used for the tails of the distributions. We jointly estimate all three HPP distributions, minimising RMSE while incorporating the RNSW transactions data and the three average holding periods. The methodology is discussed in Appendix 5. As in Section 2, the best fit was achieved with a 75-year upper bound and a combination of log-normal and beta distributions. The fitted distributions give rise to average holding periods of:

$\mu = 18.8$ years

$\mu_O = 22.6$ years

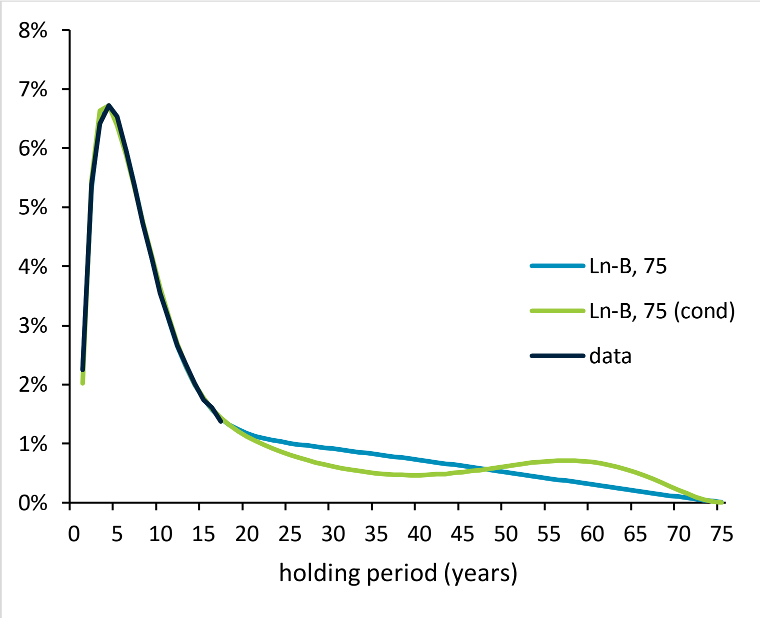
$\mu_I = 13.7$ years

In Figure 3.1 we contrast the results of the conditioned and unconditioned distributions of HPP for all purchasers. The unconditioned distribution labelled "*LN-B,75"* is the distribution identified in Section 2. The conditioned distribution, "*LN-B,75 (Cond)"* is obtained following the same methodology sketched out in Section 2 with the additional constraints on the mean of the three distributions These additional constraints result in a redistribution of the mass in the distribution beyond holding periods of 17 years, with a lower probability mass for holding periods between 17 and 47 years and an increase in the probability mass of holding periods

[1] Adjustments are based on information from the 1981 and 1986 census on the reasons why dwellings were unoccupied on census night. Rental and holiday homes are assumed not to be a principal place of residence. All other unoccupied dwellings are assumed to be split between investors and owner-occupiers in line with the shares for occupied dwellings. See SGS Economics, 'Solved: Why was no one home on census night', retrieved on 3 March 2022 from https://www.sgsep.com.au/publications/insights/why-was-no-one-home-on-census-night.

above 47 years. The displacement of probability mass towards the tail of the distribution results in an increase in the average HPP: compared to the unconditioned distribution, *LN-B,75,* the average HPP of the conditioned distribution is over 1½ years longer.

Figure 3.1: Conditioned HPP distribution – All Purchasers



	Ln-B, 75	Ln-B, 75 (cond)
mean	17.2	18.8
median	9.7	9.7
RMSE	0.180%	0.389%

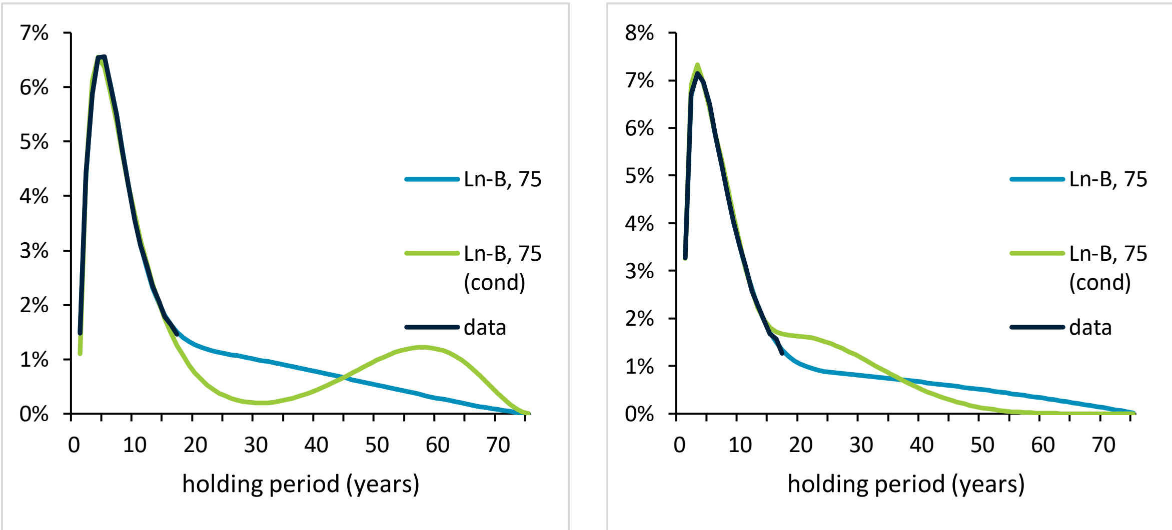
The conditioned distribution of HPP has two local maxima, with the highest peak at four years and a second local maximum at holding periods of 57 years. The probability of a newly acquired property being sold increases initially before peaking at four years. Beyond that point the probability of selling falls, suggesting that inertia builds up with length of ownership, likely related to family-lifecycle drivers, such that owners who have not sold their property within some threshold period will tend to stay there for a long time. The second local maximum is consistent with owners reaching an age where they are likely to sell to downsize or move into aged care facilities, or where there is a high probability of dying of natural causes.

In the left panel of figure 3.2 we show the results of the conditioned and unconditioned distributions of HPP for owner occupiers. Analogous results for investors are shown in the right panel of figure 3.2. The conditioned distributions are the result of a joint optimization of the three distributions (owner occupiers, investors, and all buyers), where we search for the closest fit of the estimated distributions to RNSW transactions data and the average holding periods calculated using aggregated data as explained above.

Again, all the estimated distributions are good fits for the parts of the empirical distributions that can be observed. Compared with the distributions estimated in Section 2, the conditioned HPP distribution for owner occupiers shows a shift of mass from the middle of the distribution to the right tail. The average owner-occupier holding period is 5.4 years longer in the conditioned distribution than in the distribution estimated in Section 2. The conditioned HPP distribution for investors shifts a mass from the middle of the distribution to the left part of the tail. The average investor holding period is 2.3 years shorter in the conditioned distribution than in the distribution estimated in Section 2.

Figure 3.2: Conditioned distributions, differentiated by owner type

Owner-occupiers | **Investors**



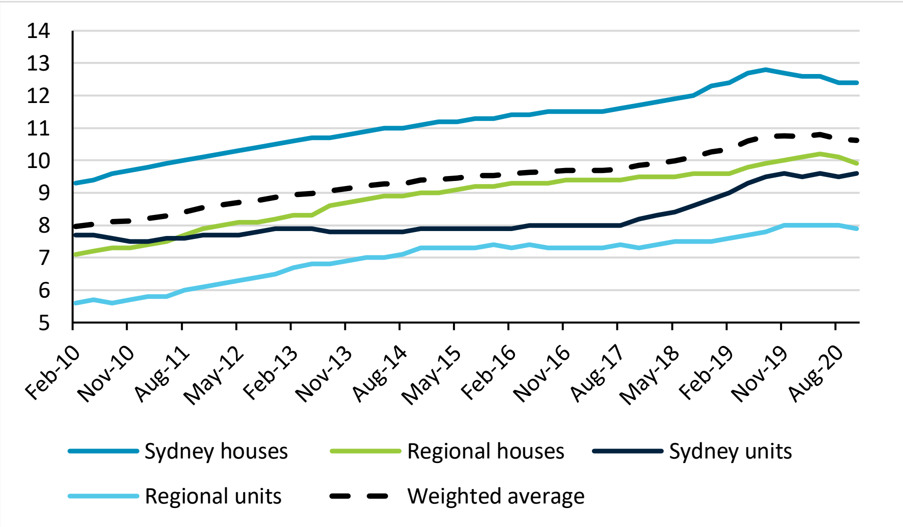
	Ln-B, 75	Ln-B, 75 (cond)
mean	17.6	22.6
median	10.3	10.5
RMSE	0.196%	0.609%

	Ln-B, 75	Ln-B, 75 (cond)
mean	16.7	13.7
median	8.9	8.8
RMSE	0.171%	0.588%

4. Property data companies' holding period estimates

Property market data specialists REA Group and CoreLogic periodically publish estimates of average holding periods. Figure 4.1 sets out data on average holding periods published by REA Group in February 2021, while Table 4.1 summarises data published by CoreLogic in September 2019. The two companies provide similar estimates, although there are some differences reflecting their respective databases and potentially some differences in methodology. We focus on the REA Group estimates.

Figure 4.1: Average holding periods in NSW – REA Estimates



Source: https://www.realestate.com.au/news/why-aussie-home-owners-are-holding-property-longer-than-a-decade-ago/ (retrieved February 2022). Average calculated by NSW Treasury using a constant weighting of regions and property types derived from Revenue NSW transactions data.

Table 4.1: Average holding periods in NSW – CoreLogic estimates

	2009		2014		2018		2019	
	House	Unit	House	Unit	House	Unit	House	Unit
Greater Sydney	9.0	7.3	10.8	8.3	11.7	8.7	12.4	9.6
Rest of NSW	7.7	6.8	9.8	8.7	10.4	9.0	10.5	9.2

Source: https://www.corelogic.com.au/news/length-home-ownership-continues-rise (retrieved, July 2021; link no longer functioning). The same information was available at https://www.smartline.com.au/mortgage-broker/jazzopardi/blog/length-of-home-ownership-continues-to-rise-2/ in February 2022.

The holding period estimates provided by REA Group differ from our estimates in two ways. First, the REA weighted average holding period ranges from 8.0 years to 10.6 years, while we estimate an average residential holding period of 18.8 years. Second, the REA averages have increased by more than 2.5 years over the course of a decade, while by construction our estimates of average holding periods are constant over time.

Three possible factors contributing to these differences are: (i) REA Group's estimates capture variations in holding periods associated with the property market cycle; (ii) the increasing trend in REA Group estimates reflects a shift in consumer behaviour; or (iii) both the lower estimates and the increasing trend over time in the REA Group data result from the absence of an adjustment for censored data.

Cyclical variations

Inspection of Figure 4.1 suggests there is a cyclical component in the REA Group estimates. For example, from mid-2017 to mid-2019, the Sydney property market saw declining volumes and declining prices. During this period, the REA Group estimates saw average Sydney house holding periods increase from 11.5 years to 12.8 years. From mid-2019, the property market recovered with significant growth in transaction volumes and resurgent price growth. By November 2020, REA Group's estimate of average Sydney house holding periods had fallen to 12.4 years. Similar cyclical variations can be seen in the data for other States recorded by REA Group.

REA Group does not publish its methodology for estimating average holding periods. We understand that they collect data on properties sold within a three-month interval, and then calculate the average time that each of those properties was owned. If the original purchase date of a property is not in the company's database, the property is not included in the calculations.

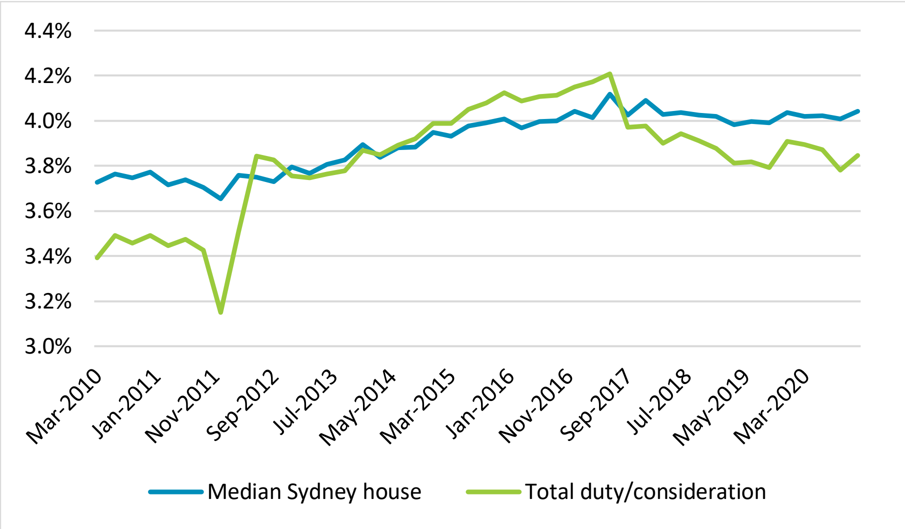
The use of three-monthly intervals means that REA Group captures cyclical movements in holding periods. In contrast, our estimates report average holding periods across 17 years of data, so that by construction no cyclical variations are reported. Nevertheless, the period mid-2017 to late 2020 captures most of a complete housing transactions cycle (peak to peak in volumes). The variations in REA Group's estimates during this period do not explain the differences in level between their estimates and our estimates, nor do cyclical variations explain the maximum variation of 2.8 years within the weighted average of REA Group's estimates.

Shift in consumer behaviour

The second possible explanation for the differences in estimates is that consumer behaviour has shifted over the course of a decade. The most obvious explanation for why holding periods might lengthen over time is that progressive rates of stamp duty, interacting with rising property prices, have resulted in an increase in average transaction costs. This in turn would cause fewer people to transact and lengthen average holding periods.

Figure 4.2 charts two measures of typical duty rates in NSW. The first is the amount of duty paid on the median Sydney house. The second divides total duty revenue by total consideration for residential transactions, as recorded by Revenue NSW. The second measure varies by more than the first, because it incorporates the effect of policy changes in the level of stamp duty concessions offered for first home buyers and others, as well as changes in the market shares of people receiving duty concessions.

Figure 4.2: Stamp duty rates in NSW



Source: Revenue NSW for duty revenue and consideration. ABS data for median house prices. NSW Treasury calculations.

During the period covered by the REA data on holding periods, (i.e., between the beginning of 2010 and the end of 2020), the average duty paid on the median Sydney house grew from 3.73 per cent to 4.04 per cent, an increase of 0.31 percentage points. Over the same period, the average duty paid on NSW residential transactions rose from 3.49 per cent to 3.85 per cent, an increase of 0.39 percentage points. To maximise the potential effect of duty on holding periods, we focus on the second measure.

Studies reviewed by Malakellis and Warlters (2021) suggest a reduction in duty rates from 4 per cent to 3 per cent will increase transaction volumes by around eight per cent. An increase in duty from 3.49 per cent to 3.85 per cent could thus lower transaction volumes by around 3.0 per cent.[2] Noting the inverse relationship between transaction volumes and holding periods, a 3.0 per cent decrease in volumes suggests the average NSW holding period could increase from 8.0 years in 2009 to 8.25 years.[3] The weighted average of REA's holding period estimates at the end of 2020 is 10.6 years.

We expect that increasing duty rates will lengthen holding periods over time. Empirically, however, changes in effective duty appear to play a relatively small role in the observed increases in REA's holding period estimates – about 0.25 years out of an observed increase of 2.6 years.

---

[2] A 0.39 percentage point reduction in duty would increase transaction volumes by approximately 0.39 x 8% = 3.12%. A duty increase of 0.39 percentage points would thus multiply transaction volumes by 1/(1.0312), which is a 3.0% reduction.

[3] An 8.0 year average holding period suggests 1/8 = 12.5% of the stock is traded each year. If this volume falls by 3%, we have 12.125% of the stock traded, suggesting an average holding period of 8.25 years.

Censored data

We believe most of the differences between REA Group's holding period estimates and our estimates arise because REA Group's methodology does not adjust for censored data. We cannot be sure, however, because REA Group's methodology is not public.

Our understanding is that REA Group's measure is based only on sales recorded in their proprietary database, and only then if a previous purchase date for a property is already in their database. REA Group's database will mature over time so that fewer and fewer transacted properties are omitted from their annual calculations because their prior purchase is not in the database. Until this occurs the holding period estimates for transactors that are derived from this data will be biased to the downside, as there will be an over-representation of properties that transact frequently. Further, as the sample of properties broadens, we should expect to see more and more long duration holding periods in the sample. This would explain why REA Group's measure of average holding periods increases over time. We also understand that CoreLogic's estimates do not adjust for censored data, which would explain why their estimates are similar to those of REA Group.

To test our hypothesis, we can suppose that the true distribution of purchaser holding periods is the conditioned distribution illustrated in Figure 3.1. If we draw observations only from the first 21 years of this distribution, the expected holding period is eight years, matching the weighted estimate of REA's estimates for NSW dwellings in February 2010. If we add an extra 11 years to the sample period, drawing from the first 32 years of the distribution, the expected holding period is 9.9 years, compared with REA weighted average of 10.6 years in November 2020. That is, the use of censored data without any adjustment could explain the difference in levels between REA Group's estimates and ours, and around 1.9 years out of the 2.6 years of growth in the weighted average of REA Group's estimates.

It is reasonable to conclude that censored data is the main reason for the difference in levels between the REA Group estimates and our estimates, and the main explanation for the increase in REA Group's estimated holding periods over the course of a decade. The same explanation potentially applies in relation to CoreLogic's estimates. It appears that neither REA Group nor CoreLogic adjust their average holding period estimates for the unsampled holding periods that are greater than the ages of their databases.

5. Surveys of time in current residence

A survey of property owners cannot provide direct evidence on purchaser holding periods, because at the time of the survey owners cannot know with certainty when they will dispose of their dwellings. Rather, a survey of owners provides evidence of 'incomplete tenure', which is the time between when a property is purchased and the time when the current owner is surveyed about how long they have owned it.

HILDA and the ABS conduct surveys that record the length of time owner-occupiers have lived in their current homes. The length of time in an owner-occupier's residence is not completely synonymous with the length of ownership, because a single owner can move between being an owner-occupier or an investor without disposing of their dwelling. Nevertheless, it would seem likely that the two periods would be closely related.

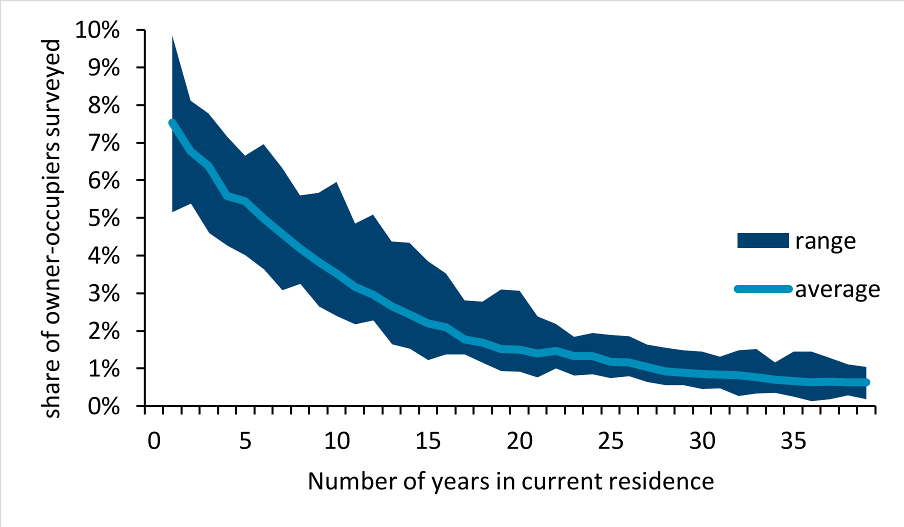
In Appendix 1 we show how an incomplete tenure distribution can be mapped into a holding period distribution, and vice versa. In this section we explore whether the surveys provide a plausible estimate of incomplete tenures, which could be used as an alternative evidence base to estimate the distribution of owner-occupier purchaser holding periods.

Description of the surveys

The HILDA survey is a longitudinal study that surveys a group of participants every year over the course of their life. Each survey is referred to as a wave. As participants enter the survey they are asked when they moved into their current home, and in subsequent surveys they are asked if they have changed their address. We treat their responses as evidence of 'residency duration'.

In Figure 5.1 we summarise data on residency duration from the HILDA survey for the years 2001 to 2018. For each of these waves, we have filtered the HILDA survey to obtain estimates for NSW owner-occupiers. The light blue line is the estimated distribution of years in current residence, with each point on the line representing the average share of respondents reporting a given residency duration across all waves. For example, on average the share of owner-occupiers that reported having been in their current residence for 15 years is 2.2 per cent. The dark blue band identifies the range of the shares across waves. For example, for residency duration of 15 years the highest reported share was 3.8 per cent in 2005 and the lowest was 1.2 per cent in 2010. The data available to us groups all residency durations of 40 years or more in a single category (which we do not report in Figure 5.1).

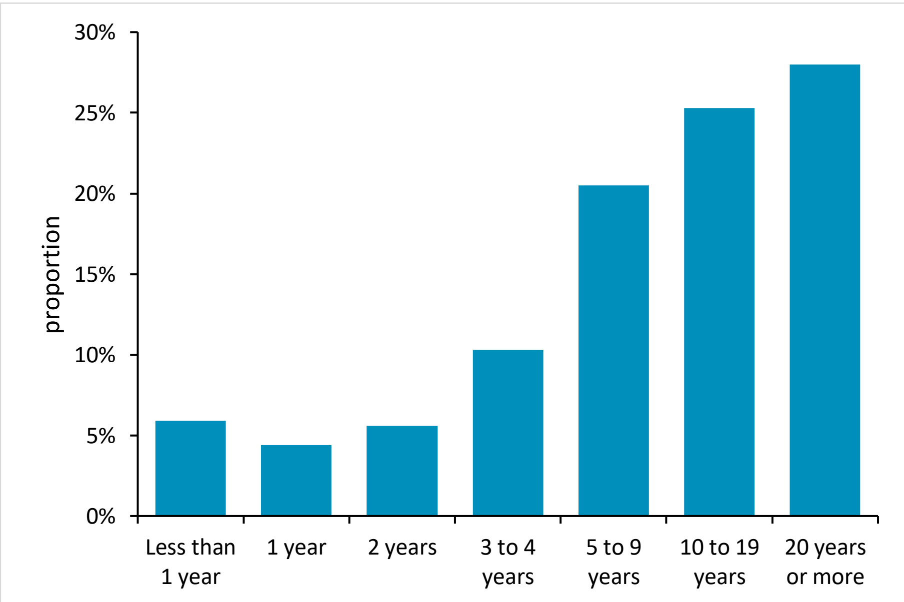
Figure 5.1: Owner-occupier years in current residence – HILDA survey data, NSW



Source: HILDA survey

The ABS, in its Survey of Housing and Income, asks a question similar to that in the HILDA survey (i.e., how long the household, which is qualified as an owner-occupier, has lived in their current residence). In figure 5.2 we show ABS data for owner-occupier residency durations in Australia for 2013-14.[4]

Figure 5.2: Years in current residence – owner-occupiers, Australia – ABS, 2013-14



Source: ABS Survey of Housing and Income 2013-14

[4] ABS statistics on residency durations by State do not distinguish between owner-occupiers and renters. The national statistics make clear that renters have significantly shorter tenures than owner-occupiers, so the State level statistics do not provide a reliable indication of residency duration for owner-occupiers.

The HILDA and ABS data yield qualitatively similar results. The median residency duration is just under 10 years in the HILDA survey and just over 10 years in the ABS survey.[5] We focus our analysis on the HILDA data.

Estimating holding periods using the HILDA data

Deriving holding period estimates from HILDA data raises various conceptual issues.

In some years, the observed values increase with an additional year of residency duration. Under the assumption of a stable holding period distribution, this is not possible. We show in Appendix 1 that the probability distribution of incomplete tenures is monotonically decreasing.

The HILDA data also present a challenge of censored data. Figure 5.2 indicates what the shape of the distribution might look like for residency durations up to 39 years. For the 18 waves of data captured in the HILDA surveys from 2001 to 2018 between six per cent and eight per cent of the observations are for durations greater than 39 years.

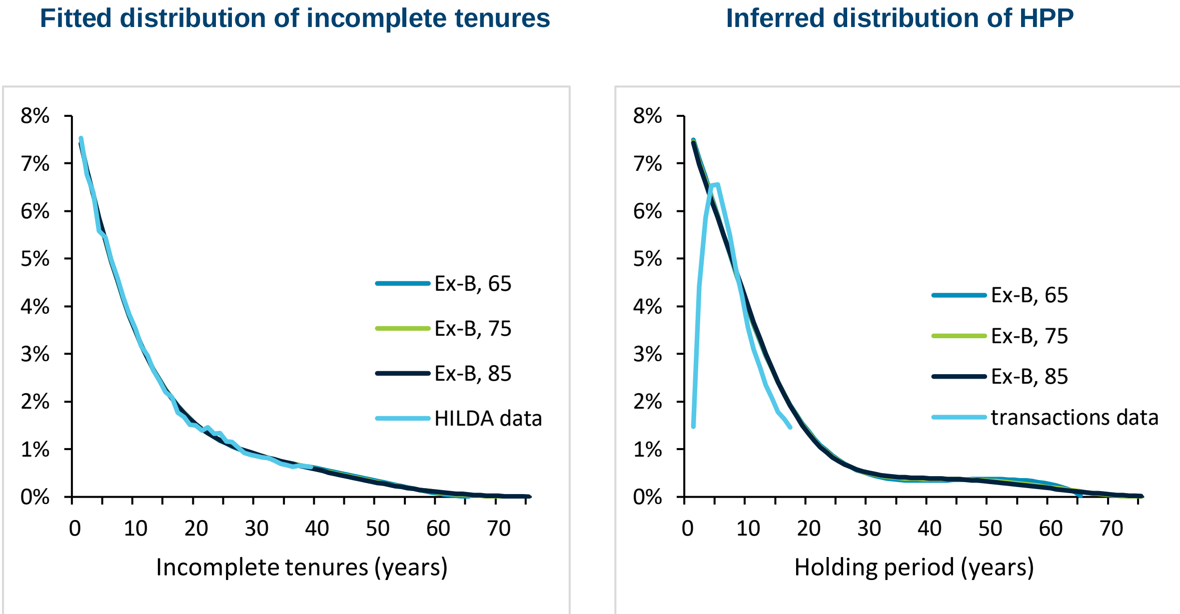
Both of these issues can be resolved by fitting an assumed functional form to the HILDA data. We tested several approaches to fitting theoretical distributions to the HILDA data. An exponential distribution fits the HILDA data moderately well, but it carries the implication that dwellings can be held for an infinite time. It can also be shown that if incomplete tenures have an exponential distribution, the implied holding period distribution must also be exponential and monotonically decreasing. That is, choosing this functional form for incomplete tenures eliminates the possibility of a local maximum within the interior of the domain of the holding period distribution, as was observed in the transaction data.

To impose an upper bound on holding periods, we fitted a combined exponential distribution and a beta distribution to the HILDA data, using the general approach to fitting theoretical distributions described in Appendix 2. We tested three alternative assumptions about the maximum holding period (65, 75 and 85 years). The best fit was obtained with an 85-year maximum holding period, although all three functions fitted the data closely. Treating the resulting functions as evidence of incomplete tenure, we calculated the distribution for the holding period of purchasers (using Equation A1.5 in Appendix 1).

The results are illustrated in Figure 5.3. The left panel shows the functions fitted to the HILDA data. The right panel shows the resulting HPP distributions. Unlike the HPP distribution derived from transactions data in Section 3, the highest likelihood of sale is in the first year of ownership, with a probability of 7.4 per cent. The estimated mean HPP is 13.5 years.

[5] Since the median falls well short of incomplete tenure lengths that are 40 years or greater these estimates are not affected by our assumption about the tail of the distribution.

Figure 5.3: Mix of Exponential and Beta distributions fitted to HILDA data



	Ex-B, 65	Ex-B, 75	Ex-B, 85
mean	13.8	13.8	13.9
median	9.2	9.2	9.2
RMSE	0.484%	0.474%	0.469%

	Ex-B, 65	Ex-B, 75	Ex-B, 85
mean	13.5	13.5	13.5
median	8.3	8.3	8.3
RMSE	1.643%	1.655%	1.668%

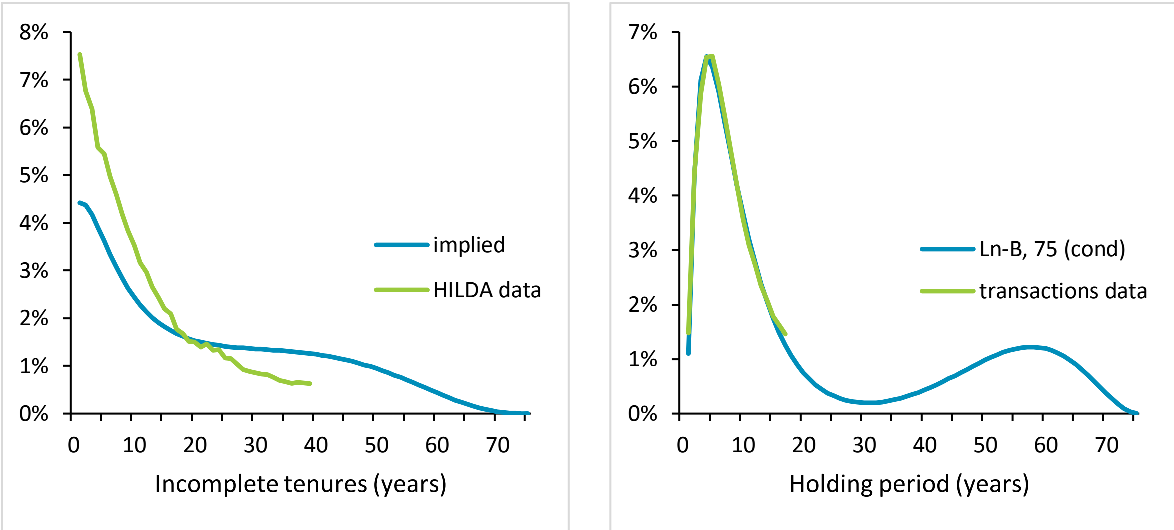
We experimented with an approach to fitting the HILDA data that could generate an increasing likelihood of sales in the early years of ownership, as observed in transactions data (i.e., a local maximum within the interior of the domain). To do this, we made assumptions about the 6 parameters that generate the combined functional forms used in Section 3 for the holding period distribution. We mapped the resulting HPP distribution into a distribution of incomplete tenures, using Equation A1.4 from Appendix 1. We then varied the starting assumptions, in order to minimise the sum of squared errors between the derived incomplete tenures and the HILDA data.

We used Excel's Solver to find the HPP parameter assumptions that generated the best fit. Depending on the seed values used as starting assumptions, we were able in some cases to reproduce the HPP global maximum observed in Section 3 at around four to six years of ownership. But in other cases, different seed values produced different shaped HPP distributions, with high probabilities of sale in the first year of ownership, or spikes at around 40-50 years. We concluded that these results did not provide a reliable basis for estimating the HPP distribution.

Finally, we supposed that the HPP distribution estimated in Section 3 is the 'true' distribution and derived the implied incomplete tenures of owner-occupiers. The results are presented in Figure 5.4. The right panel reports the assumed HPP distribution, mapped against transactions data previously set out in Section 3. The left panel compares the implied incomplete tenures with the HILDA data. The median implied incomplete tenure is 18 years, while the median duration derived from HILDA data is 9.2 years.

Figure 5.4: Incomplete tenures derived using Section 3's preferred HPP distribution

Fitted distribution of incomplete tenures | **Assumed distribution of HPP**



	implied	HILDA data
mean	22.1	N/A
median	18.0	9.2
RMSE	6.525%	N/A

	Ln-B, 75 (cond)	transactions data
mean	22.6	N/A
median	10.5	10.3
RMSE	0.609%	N/A

Reasonableness of holding period estimates derived from HILDA data

There are considerable differences between the HPP distributions estimated in Section 3 and the HPP distributions that can be derived from the HILDA data, as set out in Table 5.1.

Table 5.1: Differences between estimates of the owner occupier HPP distribution

	Section 3	Using HILDA data
Mean owner occupier HPP	22.6 years	13.5 years
Probability of sale in Year 1	1.1%	7.4%
Median incomplete tenure	18 years	9.2 years
HPP functional shape	Two internal local maxima	Monotonically decreasing

In choosing between these two approaches to estimating the HPP distribution, we place great weight on information derived from aggregate housing market data. While the housing market is subject to cycles, the long-run average share of the established stock traded each year is about 5.4 per cent. This implies that across all purchasers, the average holding period is in the order of 19 years. For life cycle and transaction cost reasons, we expect owner-occupiers to hold for longer than investors (and as expected, we observe this in the transactions data). We thus expect the mean owner-occupier holding period to be greater than 19 years. The average holding period of 13.5 years implied by HILDA data is not consistent with aggregate housing market data.

Moreover, the HILDA data appear to suggest a very high share of owner-occupiers sell within the first year of ownership. This is not consistent with observed market practice, which is influenced by the very high transaction costs of trading in dwellings.

This suggests the HILDA data are not reliable for inferring holding periods for owner occupiers. There are several reasons why this may be so.

One possible explanation is that the question asked in HILDA surveys is not about ownership, but rather it is about time spent in a currently owner-occupied dwelling. Owners may rent out their property before they move in, which would shorten their reported incomplete tenure. Similarly, owners may rent out their home after a period of owner-occupation, and these people (with longer holding periods) may not be captured in a survey of current owner-occupiers. As a result, the HILDA survey is likely to have a downward bias in reporting the period of ownership.[6]

Another difficulty is that the HILDA sample is selected to be representative of certain demographic features of the community. It is not guaranteed that this sample is representative of the distribution of holding periods across the stock of current owners.

Finally, the HILDA sample may be biased as a result of a growing stock of dwellings and a growing population of owner-occupiers. With growth, some owner-occupiers enter the survey in the lowest tenure category (i.e., less than a year for an annual survey like HILDA). Although the new owner occupiers may have the same holding period distribution as other buyers (the distribution of HPP is stable over time) their introduction into the sample will bias the average tenure down because they are not matched by households with longer tenures in the sample.

We conclude that the HILDA data cannot be used to infer the distribution of holding periods of purchasers.

[6] A related issue arises in respect of the Revenue NSW transactions data flags used to derive HPP distributions for investors and owner-occupiers in Section 3. But the transactions data accurately capture lengths of ownership, so the holding period distribution for all purchasers is likely to be accurate. The issue in using transaction data to estimate the holding periods of owner-occupiers is the weighting of ownership periods between either investors or owner-occupiers.

If the long-run investor share of the stock is stable, each transition of a dwelling owner from one state to another (e.g., from owner-occupier to investor) must on average be offset by another transition in the other direction. To the extent that transitions of usage of individual dwellings result in weighting errors, these weighting errors cancel out to some degree in the transactions data.

We cannot be sure that there is perfect cancelling out, because the investor share of the stock has been gradually increasing over time. It is also possible that among homes that transition between owner-occupancy and rental during a single period of ownership, there may be a bias toward one form of use. This could alter the weighting of holding period data for estimating owner-occupier and investor holding periods.

The transactions data must, however, have some degree of cancelling out of weighting errors. In contrast, with the HILDA data, owner-occupied duration cannot exceed the period of ownership, so occupancy can only ever be biased down as a representation of incomplete tenure. The more these transitions occur, the greater the bias.

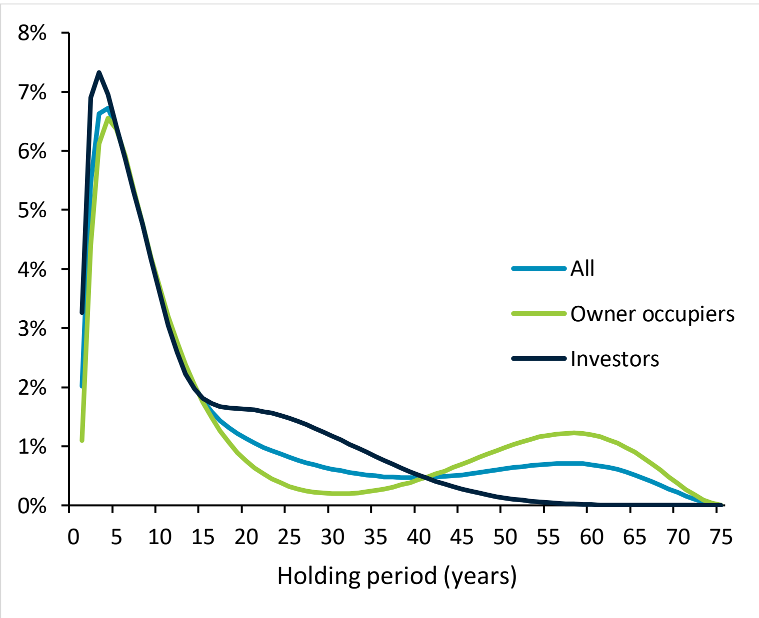
6. Conclusions

The main challenge in determining purchaser holding periods from transaction data is the limited period for which data are available, compared with the long period for which dwellings may be held. We have used Revenue NSW data which includes all purchases of residential properties from 2004-05 to 2010-11. We identified the next date of resale for each of these properties, with the most recent year of resale being 2020-21. The proportion of these properties that remain unsold by 2020-21 ranges from 29.1 per cent to 50.5 per cent, depending on the year of purchase. The data provide a firm foundation for inference about the likelihood that a dwelling will be sold within the first 17 years of ownership.

Beyond a 17-year holding period, assumptions need to be made about the distribution of holding periods. Because NSW dwellings are largely held by natural persons, the maximum holding period is limited by the duration of human lives. The share of dwellings that are unsold at the end of our data sample can be distributed across these remaining years, using different assumptions about the shape of the holding period distribution. We have combined the Revenue NSW data with data on aggregate transactions and the dwelling stock to identify the most likely distribution for the unsampled period.

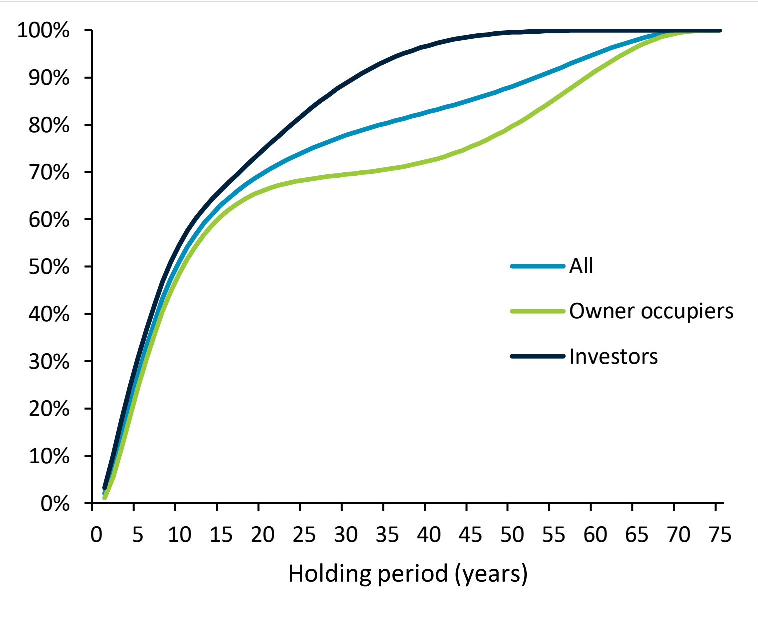
Our preferred estimates of purchaser holding periods are summarised in the distribution functions of Figures 6.1 and 6.2. Tables for the probability distribution functions and cumulative distribution functions are in Appendix 6.

Figure 6.1: Probability distribution functions



The probability distribution of holding periods of owner-occupiers has two local maxima, with the mode at four years and another local maximum at 57 years. If an owner-occupier holds for more than four years, the likelihood of sale diminishes in subsequent years, until 31 years of ownership. From this point the likelihood of sale increases again until 57 years of ownership. A picture emerges of owner-occupiers who trade frequently or hold on to their homes for much of their life. In contrast, for investors, the ownership year with the highest probability of sale is three years, and the probability of sale diminishes in all subsequent years. Lifecycle considerations seem to play a smaller role in investor transaction decisions.

Figure 6.2: Cumulative distribution functions



The cumulative distribution functions reveal that half of all purchasers sell their dwelling within 9.7 years; 60 per cent sell within 14 years; 70 per cent sell within 21 years and 80 per cent sell within 35 years. Average holding periods are 18.8 years for all purchasers, 22.6 years for owner-occupiers and 13.7 years for investors. The median holding periods are 9.7 years for all purchasers, 10.5 years for owner occupiers and 8.8 years for investors.

Caveats on the reliability of our estimates should be noted. The available sample covers a relatively short historical window compared to the maximum holding period. There is a possibility of bias in our estimates arising from cyclical variations in annual transaction volumes. Further, the proportion of censored observations increases with holding periods above 11 years, and we have no data on holding periods that exceed 17 years. Our approach to dealing with these data limitations was to fit a mixture of theoretical distributions to the available transaction data. While we considered a range of theoretical distributions our optimisations were not exhaustive in terms of the number and type of distribution and the method for mixing them.

Our estimate of the expected holding period of buyers is higher than the estimates of the average holding period of between 6 and 13 years published by property specialists such as CoreLogic and the REA Group. The estimates published by these property specialists appear to be biased down because of the way their estimates are compiled. That is, the holding period of a property that has recently sold is only calculated if that property already exists in the proprietary database. For databases that are relatively new and evolving this limitation introduces a downward bias because properties that transact frequently are likely to be over-represented.

We have also examined surveys of the number of years owner-occupiers have lived in their current homes, concentrating on the HILDA survey. If we treated years of occupancy as synonymous with years of ownership, the data would suggest the mean holding period of owner-occupiers is 13.5 years, and the median is 8.3 years. These estimates differ considerably from the estimates emerging from the Revenue NSW data and from aggregate property information on the flow of transactions and the allocation of the housing stock.

We consider that the HILDA survey does not provide a reliable basis for estimating purchaser holding periods. The discrepancy may arise because of the nature of the survey question. Years of ownership are not the same as years of occupation, because some owners may change between being an owner-occupier and being a landlord for the same dwelling. The discrepancy may also arise because the survey samples are selected to reflect certain demographic features, rather than to achieve a balanced sample of holding periods, or because samples are biased downward by the inclusion of newly built dwellings as the dwelling stock and population grow over time.

While we conclude that the HILDA survey does not provide a reliable estimate of holding periods, incomplete tenure data could be a fruitful basis for further research. In principle, a survey of the stock of owners could overcome the censored data problem. A question which directly asks a sample of owners how long they have owned their properties at the date of the survey would give an estimate of incomplete tenure, and accurate answers could be given by all respondents. With knowledge of the distribution of incomplete tenures, it would be possible to recover the distributions of purchaser holding periods, as discussed in Appendix 1. Because ownership of the dwelling stock is less volatile than the annual flow of transactions, this measure might also be less vulnerable to cyclical bias than estimates based on transaction data.

Appendix 1: Mathematical relationships between holding period measures

A holding period is the time between when a dwelling is purchased and when it is next sold, or otherwise transferred to a new owner. Despite the seeming simplicity of this idea, we have identified three distinct measures that could be used to report holding periods: purchasers' holding periods, owners' holding periods, and owners' incomplete tenures. This appendix explores the relationship between these three measures, under the assumption that the distribution of purchaser holding periods is stable over time.

(a) Overview

The holding period of purchasers (HPP) can be thought of as a random variable, with probability P_x that the dwelling will be sold in the x-th year of ownership. The average purchaser's holding period is given by $\mu = \sum_{x=1}^{B} xP_x$, where B is a finite upper bound on the length of time a person could own a property. It can be shown that the share of established properties traded each year will average, over time, $\frac{1}{\mu}$. For example, if five per cent of established properties are traded each year, the average buyer's holding period is 20 years.

The average holding period of owners (HPO) measures holding periods weighted by the stock of owners. The share of owners with a holding period of x years can be denoted S_x. The mean owner's holding period is given by $\varphi = \sum_{x=1}^{B} xS_x$. It can be shown that the mean owner's holding period is always greater than the mean purchaser's holding period: $\varphi \geq \mu$.

Finally, a survey of current owners can be conducted, asking them how long they have owned their property. At the time people respond to this question they do not know their holding period. Their response provides evidence of the 'incomplete tenure' of owners (ITO), which is the time between when a property is purchased and the time when the current owner is surveyed about how long they have owned it. The share of the dwelling stock held by owners with an incomplete tenure of x years is denoted by T_x. The mean incomplete tenure is given by: $\tau = \sum_{x=1}^{B} xT_x$. It can be shown that the mean incomplete tenure is half the mean owner's holding period: $\tau = \frac{\varphi}{2}$.

There is a one-to-one relationship between the three holding period measures: P_x, S_x and T_x. If the full distribution of one measure is known, it is possible to recover the remaining two distributions.

(b) Inverse relationship: When all buyers are drawn from the same buyer holding period distribution, the expected probability, Q_y that a property is traded in year y is the inverse of the mean buyer's holding period, μ.

$$E[Q_y] = \frac{1}{\mu}$$

After a property is bought, it will be held for an expected period of μ years. We thus expect $\frac{1}{\mu}$ transactions per property per year.

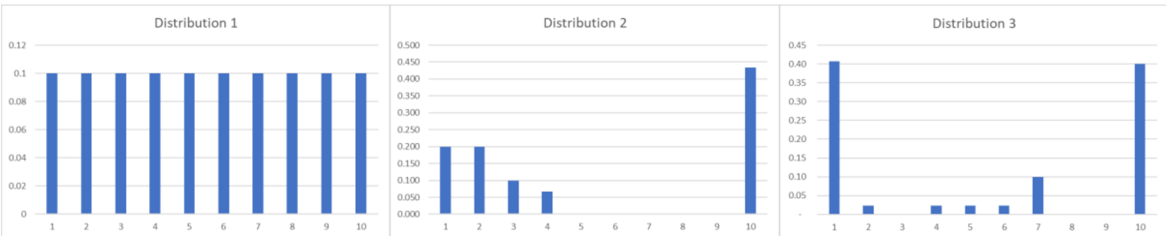
The relationship can also be specified as the probability that a property is sold in year n, as $n \to \infty$. That is:

$$\lim_{y \to \infty} Q_y = \frac{1}{\mu}$$

An example illustrates how the probability of sale of a property converges to $\frac{1}{\mu}$ as time advances, and also illustrates that many different purchaser holding period distributions are consistent with the same probability of sale.

Let purchasers be drawn from a purchaser holding period with an upper bound of 10 years. The probability that a property is sold in the n-th year after purchase is denoted P_n so that $\sum_{n=1}^{10} P_n = 1$. Three such distributions are given below. All three of these purchaser distributions have a mean holding period of 5.5 years.

	Distribution 1	Distribution 2	Distribution 3
P_1	0.1	0.200	0.40765
P_2	0.1	0.200	0.02306
P_3	0.1	0.100	0.00010
P_4	0.1	0.067	0.02306
P_5	0.1	0.000	0.02306
P_6	0.1	0.000	0.02306
P_7	0.1	0.000	0.10000
P_8	0.1	0.000	0
P_9	0.1	0.000	0
P_{10}	0.1	0.433	0.40000
Mean holding period	5.5	5.5	5.5
Median holding period	5.5	3	6



Suppose a dwelling is purchased in year zero ($Q_0 = 1$). Knowing the holding period distribution for purchasers, the probability Q_y that the dwelling is sold in any subsequent year can be expressed as:

$$Q_1 = P_1$$

$$Q_2 = P_1 Q_1 + P_2$$

$$Q_3 = P_2 Q_1 + P_1 Q_2 + P_3$$

$$Q_4 = P_3 Q_1 + P_2 Q_2 + P_1 Q_3 + P_4$$

$$\vdots$$

$$Q_{10} = \sum_{i=1}^{10} P_i Q_{10-i}$$

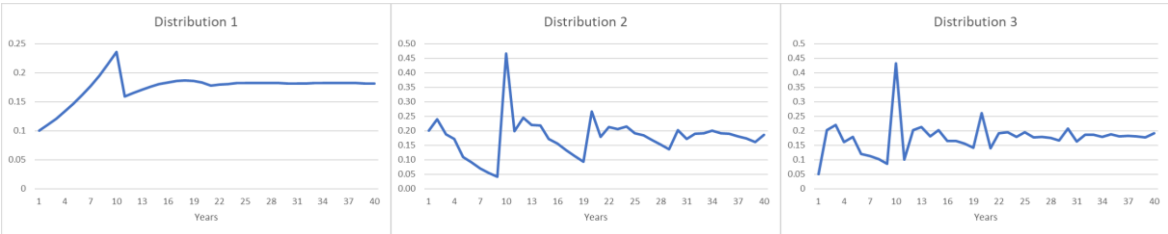
$$\vdots$$

And for all $y > 10$,

$$Q_y = \sum_{i=1}^{10} P_i Q_{y-i} \qquad \text{(A1.1)}$$

The following charts report Q_y for years one to 40, for each of the three distributions set out above. In the initial years, the probability of sale of the property in a particular year is largely driven by the first buyer's holding period distribution. As time advances, successive purchases result in a combination of multiple buyers' holding periods, and in all three cases the probability of sale of the property in year y converges to:

$$\lim_{y \to \infty} Q_y = 0.181818 = \frac{1}{5.5} = \frac{1}{\mu}$$



It is possible to sketch a proof of the inverse relationship as a limit. Generalising from equation (A1.1), the probability that a property is sold in any year y depends on the probability of sale in all preceding years, so that as $y \to \infty$ the value of Q_y is the sum of an infinite series of positive terms.

Q_y is a probability, and so has a maximum value of one. In order for an infinite sum of positive terms to be less than or equal to one, it must be true that Q_y converges, so there exists a constant $\bar{Q} = \lim_{y \to \infty} Q_y$ where $0 \leq \bar{Q} \leq 1$.

As this limit is approached, the probability that a dwelling is sold in a random year approaches a constant. That is, although buyers may have a varying probability of sale across years of ownership, the probability of sale of a dwelling converges in the long run to a constant.

Given a constant probability each year that a dwelling is sold, $\bar{Q}$, the number of years (Y) a dwelling is held is a random variable with a geometric distribution. For a geometric distribution, the expected value is:

$$E[Y] = \frac{1}{\bar{Q}}$$

But the expected number of years between transactions for a property must be equal to the expected number of years that a buyer holds a property: i.e., $E[Y] = \mu$. This in turn implies $\bar{Q} = \frac{1}{\mu}$, and so:

$$\lim_{y \to \infty} Q_y = \bar{Q} = \frac{1}{\mu}$$

(c) The inverse relationship also holds when there are two (or more) groups of buyers, each with their own distinct buyers' holding period distributions. The probability that a property is sold in a given year converges to the inverse of the average buyer's holding period, where averages are weighted by the transaction shares of each buyer group.

Suppose there are two groups of buyers—owner-occupiers and investors (O and I)—each with maximum holding periods of B years. The two groups are drawn from two distinct buyer holding period probability distributions. For the two groups, the discrete probability that the owner sells in the i-th year of ownership is given by P_i^O and P_i^I, respectively. The mean holding period for group O is given by μ^O and the mean holding period for group I is given by μ^I.

Further suppose that when a property is sold it is purchased by a person from group I with probability α or a person from group O with probability $1 - \alpha$.

If a dwelling is bought at time 0 by a random buyer, the probability that it is sold in Year y>0 is given by the recursive expression:

$$Q_y = [(1-\alpha)P_1^O + \alpha P_1^I]Q_{y-1} + [(1-\alpha)P_2^O + \alpha P_2^I]Q_{y-2} + \cdots + [(1-\alpha)P_B^O + \alpha P_B^I]Q_{y-B}$$

This expression takes the same functional form as the earlier expression (A1.1), replacing the purchaser's probability of sale in the n-th year after a purchase with a weighted average probability. It follows that:

$$\lim_{y \to \infty} Q_y = \frac{1}{\mu}$$

where the average buyer holding period is weighted by transaction shares of each buyer type:

$$\mu = \sum_{i=1}^{B} i[(1-\alpha)P_i^O + \alpha P_i^I] = (1-\alpha)\mu^O + \alpha\mu^I$$

(d) The inverse relationship between the average buyer's holding period and the long-run probability of sale of a property in a random year is observed even when there are particular groups of properties that are bought only by particular types of buyer.

For example, suppose there are nine properties whose buyers hold for one year only before selling, and one property which is always bought by a buyer who holds for exactly 10 years before selling.

For the 10 properties, over the course of 10 years there are 91 transactions. So the probability of sale of a randomly selected property in a random year is 91/100.

For the buyers of the group of nine properties, the average purchaser holding period is one year. For the tenth property the average purchaser holding period is 10 years. Across all buyers, the average purchaser holding period, weighted by transaction shares, is 90/91 x 1 + 1/91 x 10 = 100/91 years.

Again, the probability that a property is sold in a given year is the inverse of the average purchaser's holding period, where the two groups of buyers are weighted by transaction shares.

(e) In equilibrium there is a one-to-one relationship between the distribution of buyer holding periods and the distribution of owner holding periods (i.e., shares of the dwelling stock held by owners with different holding periods).

Every time a property is bought, it is bought by a buyer drawn from a discrete holding period distribution where the probability that the buyer holds for x years is given by P_x and $\sum_{x=1}^{B} P_x = 1$.

Every dwelling purchase results in one property being held for an expected period of $\mu = \sum_{x=1}^{B} xP_x$ years. The expected value is the probability weighted contribution of buyers with holding periods from one up to B years. Because the weights of buyers with different holding periods will be reproduced with each new purchase, the expected share of the dwelling stock owned by people with a buyer holding period of x years is

$$S_x = \frac{xP_x}{\mu} \qquad \text{(A1.2)}$$

That is, if we know the values of P_x for all x, we can determine the values of S_x for all x.

We can also recover the buyer holding period distribution from data about the ownership of the dwelling stock. Suppose we know the holding periods of current owners of the dwelling stock, where the share of people who sell in the x-th year of ownership is given by S_x, and $\sum_{x=1}^{B} S_x = 1$.

From equation (A1.2), we have:

$$P_x = \mu \frac{S_x}{x}$$

Accordingly,

$$\sum_{i=1}^{B} P_i = \sum_{i=1}^{B} \mu \frac{S_i}{i}$$

$$1 = \mu \sum_{i=1}^{B} \frac{S_i}{i}$$

$$\mu = \frac{1}{\sum_{i=1}^{B} \frac{S_i}{i}}$$

And so:

$$P_x = \mu \frac{S_x}{x} = \frac{S_x/x}{\sum_{i=1}^{B}(S_i/i)} \qquad \text{(A1.3)}$$

That is, if we know the values of S_x for all x, we can determine the values of P_x for all x.

(f) The expected share of the established dwelling stock traded each year is the inverse of the mean buyer holding period.

Ignoring cyclical movements in transaction volumes, among owners with a holding period of x years the expected share who sell each year is $\frac{1}{x}$. Summing across all types of current owners of the dwelling stock with different holding periods and using Equation (A1.2), the expected share of the dwelling stock traded each year is

$$\sum_{x=1}^{B} \frac{1}{x} S_x = \sum_{x=1}^{B} \frac{1}{x} \frac{x P_x}{\mu} = \frac{1}{\mu} \sum_{x=1}^{B} P_x = \frac{1}{\mu}$$

This finding is closely related to relationship (b), above. As noted in the discussion of (b), there are many buyer holding period distributions with the same mean, μ. The share of the dwelling stock that is traded each year identifies the (inverse of the) mean buyer holding period, but it does not identify the entire distribution or any other moments of the distribution such as the median.

The relationship specifies the expected holding period for the stock of owners at a point in time. To estimate the mean buyer holding period by measuring the share of the stock traded each year, sales of newly built properties should be excluded.

(g) There is a one-to-one relationship between the distribution of the holding period of owners (HPO) and the distribution of the incomplete tenure of owners (ITO).

Incomplete tenure is the number of years current owners have owned their property. Because a current owner has not yet sold their property, an owner's incomplete tenure must be less than the owner's complete holding period.

The relationship between distributions is best introduced with an example. Suppose 10% of the stock of properties are owned by people with a holding period of one year, 60 per cent of properties are held by people with a holding period of two years, and 30 per cent of properties are owned by people with a holding period of three years. That is, $S_1 = 0.1$, $S_2 =$

0.6, and $S_3 = 0.3$. The following table maps the ownership shares to incomplete tenure shares:

	Years	Incomplete tenure 1	Incomplete tenure 2	Incomplete tenure 3	Share of owners, by holding period
Holding period	1	0.1			10%
	2	0.3	0.3		60%
	3	0.1	0.1	0.1	30%
Share of owners, by incomplete tenure		50%	40%	10%	

To understand the table, note that in expectation:

- Among the 10 per cent of owners with a holding period of one year, all report an incomplete tenure of one year.
- Among the 60 per cent of owners with a holding period of two years, half report an incomplete tenure of one year, and half report an incomplete tenure of two years.
- Among the 30 per cent of owners with a holding period of three years, one third report an incomplete tenure of one year; one third report an incomplete tenure of two years; and one third report an incomplete tenure of three years.

It follows that:

- the share of all owners with an incomplete tenure of one year is $S_1 + \frac{1}{2}S_2 + \frac{1}{3}S_3 = 50\%$.
- the share of all owners with an incomplete tenure of two years is $\frac{1}{2}S_2 + \frac{1}{3}S_3 = 40\%$.
- the share of all owners with an incomplete tenure of three years is $\frac{1}{3}S_3 = 10\%$.

In the general case, the share of all current owners with incomplete tenure of x years is given by:

$$T_x = \sum_{j=x}^{B} \frac{S_j}{j} = \sum_{j=x}^{B} \frac{P_j}{\mu} \quad \text{(A1.4)}$$

That is, if we know the values of P_x for all x, or if we know the values of S_x for all x, we can determine the values of T_x for all x.

We can recover the distribution of S_x if we know the distribution of T_x with the following relationship (where $T_{B+1} = 0$):

$$S_x = x(T_x - T_{x+1})$$

We can also use the previously observed relationship between the buyer and owner holding period distributions (Equation A1.3) to recover the distribution of buyer holding periods from the distribution of incomplete tenures.

$$P_x = \frac{S_x/x}{\sum_{i=1}^{B}(S_i/i)} = \frac{T_x - T_{x+1}}{T_1 - T_{B+1}} = \frac{T_x - T_{x+1}}{T_1} \quad \text{(A1.5)}$$

(h) The probability distribution of incomplete tenures decreases monotonically (i.e., it is downward sloping).

This proposition states that the share of owners with incomplete tenures of x years is always greater than or equal to the share of owners with incomplete tenures of $x + 1$ years. It is true whatever the form of the distribution of purchaser holding periods.

Using equation A1.4,

$$T_x - T_{x+1} = \sum_{j=x}^{B} \frac{P_j}{\mu} - \sum_{j=x+1}^{B} \frac{P_j}{\mu} = \frac{P_x}{\mu} \geq 0$$

(i) Mean values

As before, assume the holding periods of all buyers are drawn from a common discrete distribution, where the probability that a buyer will hold for x years is given by P_x. The share of owners with a holding period of x years is given by S_x. The share of current owners who have lived in their property for x years is given by T_x.

The mean buyer holding period is

$$\mu = \sum_{x=1}^{B} x P_x$$

The mean owner holding period is

$$\varphi = \sum_{x=1}^{B} x S_x = \sum_{x=1}^{B} \frac{x^2 P_x}{\mu} = \frac{1}{\mu} \sum_{x=1}^{B} x^2 P_x$$

The mean incomplete tenure is

$$\tau = \sum_{x=1}^{B} x T_x = \sum_{x=1}^{B} x \sum_{j=x}^{B} \frac{S_j}{j} = \frac{1}{\mu} \sum_{x=1}^{B} x \sum_{j=x}^{B} P_j$$

(j) The mean owner holding period is greater than the mean buyer holding period

By definition of the mean owner holding period:

$$\varphi = \sum_{x=1}^{B} x S_x = \sum_{x=1}^{B} \frac{x^2 P_x}{\mu}$$

Rearranging, we have:

$$\varphi \mu = \sum_{x=1}^{B} x^2 P_x$$

By Jensen's Inequality, $E[X^2] \geq (E[X])^2$, so:

$$\varphi\mu = \sum_{x=1}^{B} x^2 P_x \geq \left(\sum_{x=1}^{B} x P_x\right)^2 = \mu^2$$

And thus:

$$\varphi \geq \mu$$

This relationship states that the average holding period when holding periods are weighted by the stock of owners is greater than the average holding period when weighted by purchasers.

If, for example, all purchasers were equally likely to hold for either one or two years (i.e. $\mu = 1 \times 0.5 + 2 \times 0.5 = 1.5$ years), by virtue of longer holding periods, 2-year holders would own twice as much of the stock as 1-year holders, so that $\varphi = 1 \times \frac{1}{3} + 2 \times \frac{2}{3} = 1.67$ years. That is, $\varphi \geq \mu$.

(k) The mean incomplete tenure is half the mean owner holding period

To establish this relationship, it is convenient to shift from discrete distributions to continuous distributions.

The mean owner holding period is:

$$\varphi = \int_0^B x S_x \, dx$$

The mean incomplete tenure is:

$$\tau = \int_0^B x T_x \, dx = \int_0^B x \int_x^B \frac{S_j}{j} \, dj \, dx$$

Reversing the order of integration gives:

$$\tau = \int_0^B \frac{S_j}{j} \int_0^j x \, dx \, dj$$

Note that:

$$\int_0^j x \, dx = \left.\frac{x^2}{2}\right|_{x=0}^{x=j} = \frac{j^2}{2}$$

so that:

$$\tau = \int_0^B \frac{S_j}{j} \frac{j^2}{2} dj = \frac{1}{2} \int_0^B j \, S_j \, dj = \frac{1}{2} \varphi$$

The intuition for this result is that when a dwelling owner is surveyed about how long they have owned their dwelling, on average the survey question will be asked halfway through the owner's holding period.[7]

[7] The literature on the duration of unemployment deals with an analogous issue in relation to incomplete unemployment durations captured in labour force surveys. See, for example, Corak and Heisz (1995).

Appendix 2: Algorithm for fitting distributions

This appendix describes the method we use to fit a theoretical distribution of holding periods to the data we have sourced from Revenue NSW.

Balancing the goals of fitting the data well while keeping our approach simple, we consider mixtures of two theoretical distributions as our candidates to match the data. Inspection of the data from Revenue NSW indicated that a single theoretical distribution was unlikely to be sufficient. Also, single distributions tended to fare poorly when we tried to fit both Revenue NSW data and the average holding periods calculated using aggregate data. For this exercise, which we discuss in more detail in Section 3 and Appendix 5, a mix of two distributions was essential to produce a satisfactory match—with one distribution used to fit the data from Revenue NSW and the other distribution used to shape the right tail of the mixed distribution and match the average holding periods based on aggregate data.

An additional consideration that has guided our approach is that holding periods must be finite and related to the natural life span of owners. This means that the distribution that we fit to the data must be bounded.

We maintain tractability by narrowing our choice to two-parameter distributions. The distribution shaping the left side of the mixed distribution can be drawn from Weibull, lognormal or beta distributions. We refer to this distribution within the mix as the primary distribution. The main purpose of the primary distribution is to fit the part of the empirical distribution of holding periods revealed by the Revenue NSW vintages. The distribution shaping the right side of the mixed distribution is drawn from the beta distribution, which is bounded. We refer to this distribution as the secondary distribution.

To piece the primary and secondary distributions together, we use the cumulative distribution function (CDF) of a beta distribution to shave the right tail of the primary distribution. We refer to this as the truncation function. The secondary distribution is scaled to capture the mass of the primary distribution that has been shaved off. Together, the truncated primary distribution and the scaled secondary distribution add up to the candidate mixed distribution.

An optimisation routine is used to jointly choose the parameters of the primary and secondary distributions and the truncation function to minimise the root mean square error (RMSE) between the overlapping parts of the mixed distribution and its empirical counterpart. Below we set out the details of this optimisation.

Optimization algorithm

Let h denote holding periods.

Step 1: Select the maximum holding period.

— Let H denote our assumption for the upper bound of the holding periods distribution. We tested upper bounds of 65, 75, and 85 years.

Step 2: Select the family of theoretical distributions for the primary distribution, the secondary distribution, and the truncation function.

— The *Primary* distribution, $PR(h, \alpha_1, \alpha_2)$, is a two-parameter distribution used primarily for fitting the empirical data from Revenue NSW. The candidate distributions are lognormal, beta, and Weibull.

— The *Secondary* distribution $SD(h, \alpha_3, \alpha_4)$, is a 2two-parameter, bounded distribution that allows us to shape the right tail of the mixed distribution. A beta distribution is used for this purpose.

— The *Truncation* function, $T(h, \alpha_5, \alpha_6)$, is a two-parameter distribution used to assign weights to the observations drawn from the primary distribution. If these weights are all equal to 1 then we recover the primary distribution without truncation. Weights less than 1 will introduce truncation into the primary distribution. The truncation function is the CDF of a beta distribution, which is a monotonically increasing function that ensures a smooth truncation.

Step 3: Construct a mixed distribution.

The primary and secondary distributions and the truncation functions are used to construct the mixed distribution $F(h, \bar{\alpha})$:

$$F(h, \bar{\alpha}) = f\big(PR(h, \alpha_1, \alpha_2), SD(h, \alpha_3, \alpha_4), T(h, \alpha_5, \alpha_6)\big)$$

where $\bar{\alpha} = \{\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6\}$ and $f(\cdot)$ denotes the method we use to calculate the mixed distribution based on the primary and secondary distributions and the truncation function. Details of this calculation are discussed in the example below.

Step 4: Solve an optimisation problem.

— The choice variables are the six parameters that characterise the primary and secondary distributions and the truncation function, $\bar{\alpha} = \{\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6\}$.

— The objective function is the root mean squared error (RMSE) between the overlapping parts of the mixed distribution and $Q(h)$, the empirical distribution we recovered from Revenue NSW transactions data. That is:

$$RMSE(h, \bar{\alpha}) = \sqrt{\frac{1}{\bar{h}}\sum_{h=1}^{\bar{h}} \big(F(h, \bar{\alpha}) - Q(h)\big)^2}$$

where $\bar{h}$ is the maximum holding period for which data is available (equal to 17 years for most of the exercises we run).

— Use solver in MS-Excel to search for the values of the six parameters that minimize the RMSE, i.e., that result is the smallest differences between the candidate mixed distribution and the empirical distribution.

Illustrative example

The following example demonstrates how the steps set out above are applied in practice.

Step 1: Select the maximum holding period.

— We assume for the example that holding periods cannot be greater than 75 years.

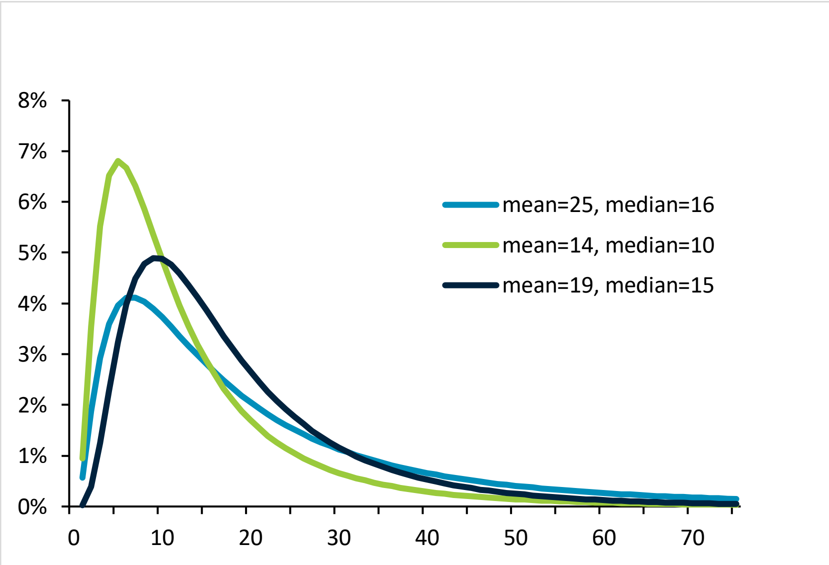
Step 2: Select the family of theoretical distributions for the primary distribution, the secondary distribution, and the truncation function.

— *Primary* distribution, $PR(h, \alpha_1, \alpha_2)$. Suppose it is log-normal:

$$PR \sim \ln_p(\alpha_1, \alpha_2) \quad \text{(A2.1)}$$

Examples of lognormal distributions with different parameter values are shown in Figure A2.1.

Figure A2.1: Examples of log-normal distributions

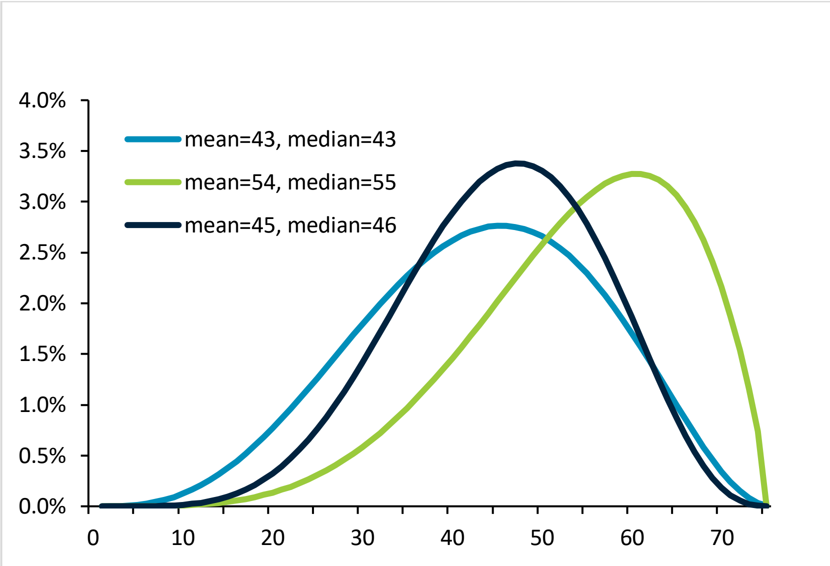


— *Secondary* distribution $SD(h, \alpha_3, \alpha_4)$. Suppose it is a beta distribution:

$$SD \sim \beta(\alpha_3, \alpha_4) \quad \text{(A2.2)}$$

Examples of beta distributions with different parameter values are shown in Figure A2.2.

Figure A2.2: Examples of beta distributions

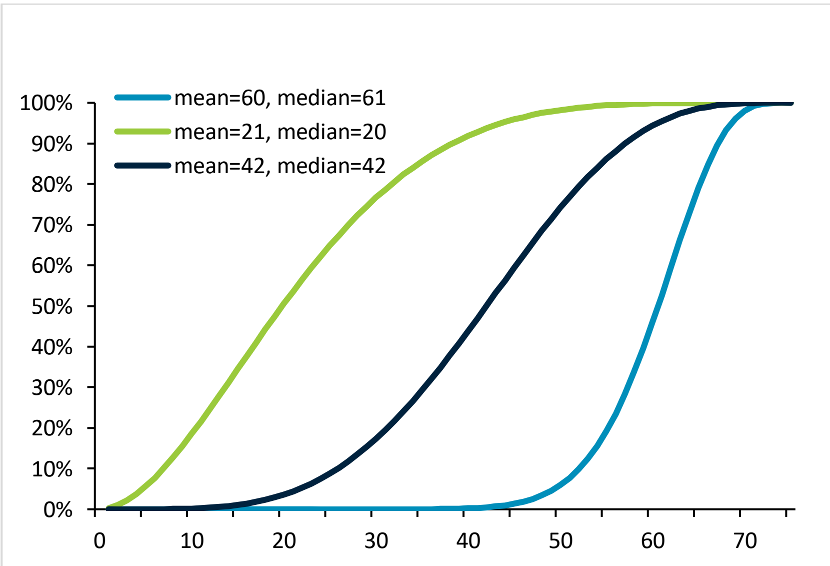


— *Truncation* function, $T(h, \alpha_5, \alpha_6)$. To mix the primary and secondary distributions we need a truncation function. We assume that this truncation functions takes the form of the CDF of a beta distribution:

$$T \sim \mathrm{B}(\alpha_3, \alpha_4) \tag{A2.3}$$

A selection of truncation functions based on beta distributions with different parameters are shown in Figure A2.3.

Figure A2.3: Examples of truncation functions



Step 3: Construct a mixed distribution.

Seed values for the six parameters that characterise the primary and secondary distributions and the truncation function are provided:

$$\bar{\alpha}^0 = \{\alpha_1^0, \alpha_2^0, \alpha_3^0, \alpha_4^0, \alpha_5^0, \alpha_6^0\} \qquad \text{(A2.4)}$$

Suppose this yields the primary distribution shown in figure A2.4 and the truncation function shown in figure A2.5.

Figure A2.4: Initial primary distribution based on seed values

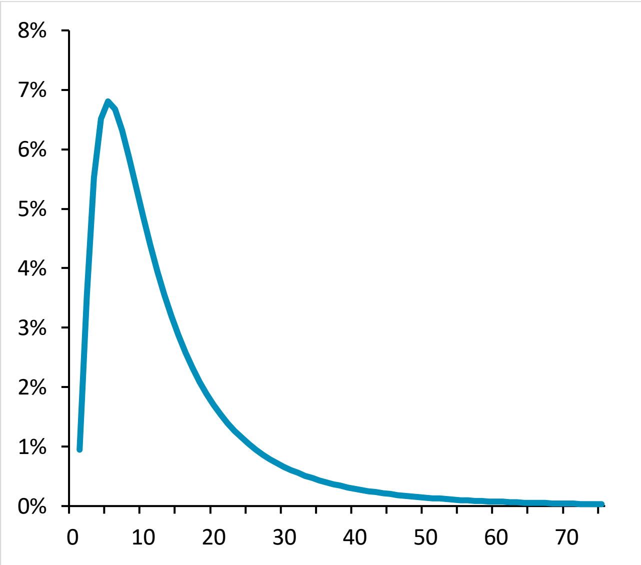
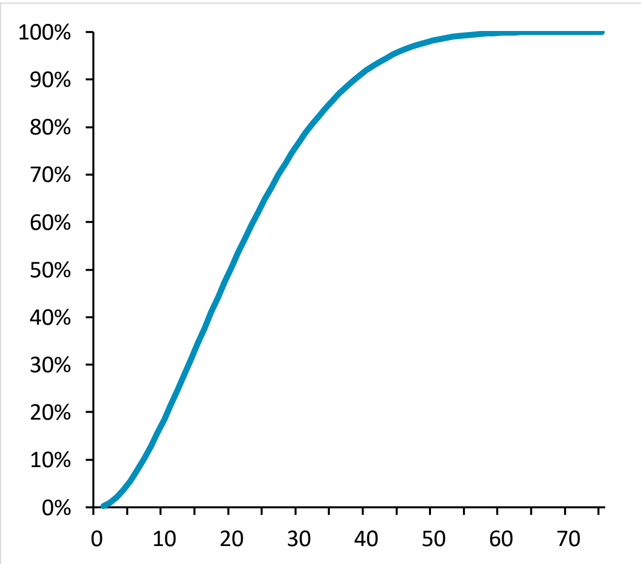


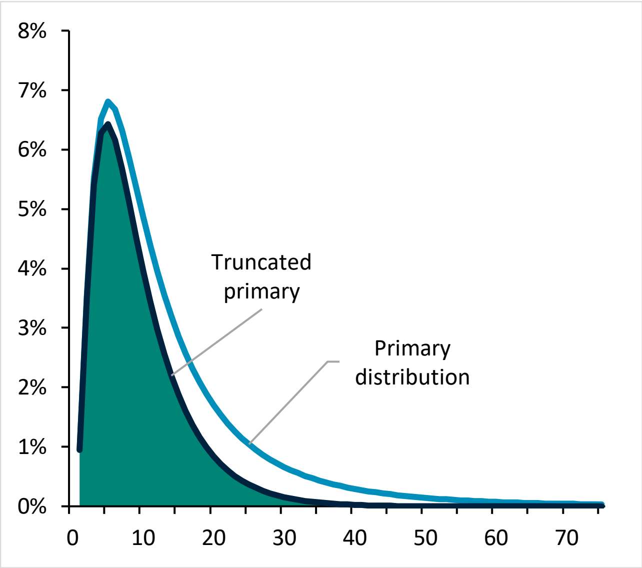
Figure A2.5: Initial truncation function based on seed values



The resultant truncated primary distribution is superimposed in figure A2.6 on the primary distribution.

$$TR(h, \alpha_1^0, \alpha_2^0, \alpha_5^0, \alpha_6^0) = PR(h, \alpha_1^0, \alpha_2^0) \times \left(1 - T(h, \alpha_5^0, \alpha_6^0)\right) \qquad \text{(A2.5)}$$

Figure A2.6: Truncated primary distribution



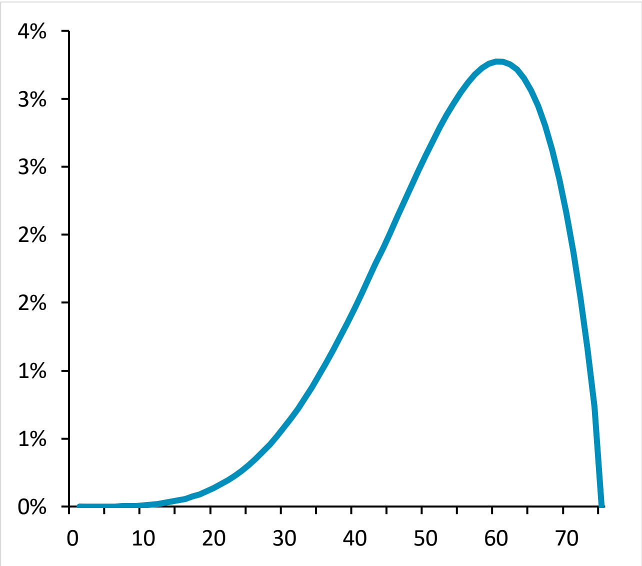
The area under the truncated primary distribution is equal to:

$$\mathrm{A} = \sum_{h=0}^{H} TR(h, \alpha_1^0, \alpha_2^0, \alpha_5^0, \alpha_6^0) \qquad \text{(A2.6)}$$

where H is 75 in this example.

The secondary distribution yielded by the initial seed vector, $\bar{\alpha}^0$ is shown in figure A2.7.

Figure A2.7: Secondary distribution

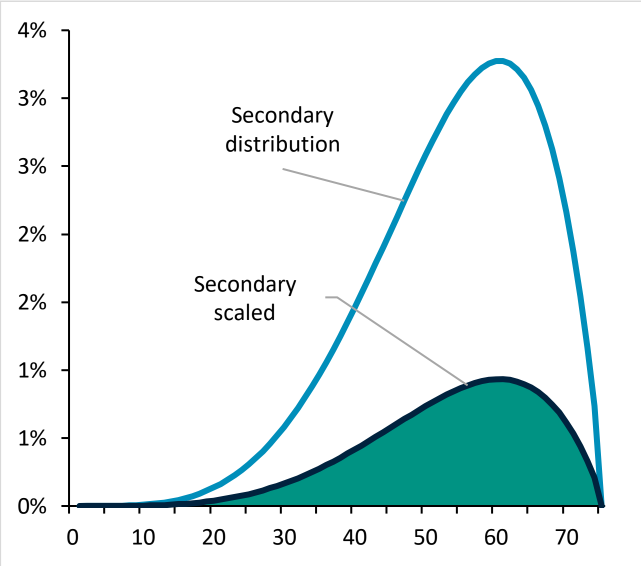


The mass in the secondary distribution is required to be equal to the mass taken out of the primary distribution. We do this by scaling the secondary distribution:

$$SS(h, \bar{\alpha}^0) = SD(h, \alpha_3^0, \alpha_4^0) \times (1 - A_0) \quad \text{(A2.7)}$$

The scaled secondary distribution is shown in figure A2.8.

Figure A2.8: Secondary distribution - scaled



Note that the area below the scaled secondary distribution is equal to:

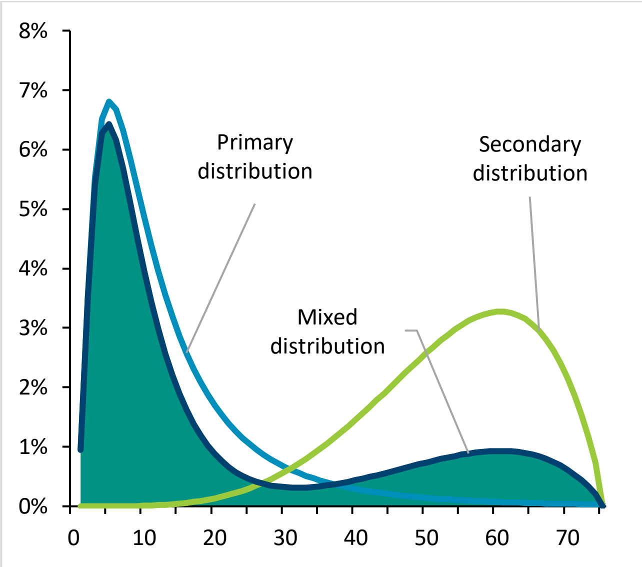
$$\sum_{h=0}^{H} SS(h, \bar{\alpha}^0) = 1 - A_0 \quad \text{(A2.8)}$$

The mixed distribution can now be calculated as follows:

$$F(h, \bar{\alpha}^0) = TR\left(h, \alpha_1^0, \alpha_2^0, \alpha_5^0, \alpha_6^0\right) + SS(h, \bar{\alpha}) \quad \text{(A2.9)}$$

This is shown in figure A2.9.

Figure A2.9: Mixed distribution



The area below the mixed distribution, represented in figure A2.9 by the green shading, is equal to:

$$\sum_{h=0}^{H} F(h, \bar{\alpha}^0) = \sum_{h=0}^{H} TR\left(h, \alpha_1^0, \alpha_2^0, \alpha_5^0, \alpha_6^0\right) + \sum_{h=0}^{H} SS(h, \bar{\alpha}) = 1 \qquad \text{(A2.10)}$$

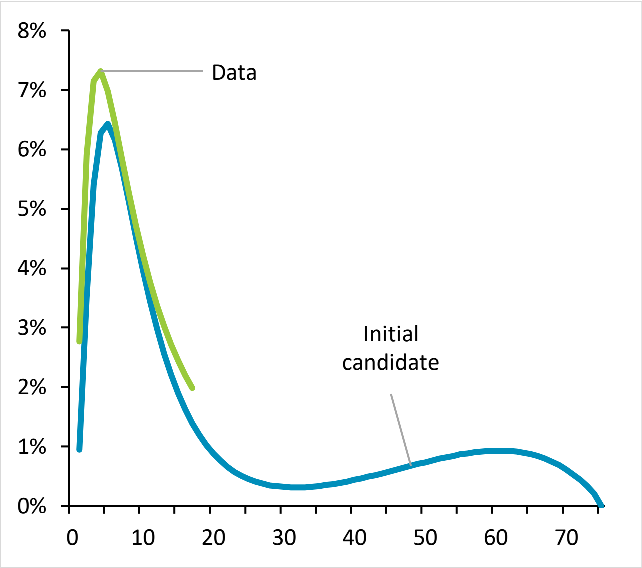
Step 4: Solve an optimisation problem.

The RMSE can now be calculated as:

$$RMSE^0 = \sqrt{\frac{1}{\bar{\bar{h}}} \sum_{h=1}^{\bar{h}} \left(F(h, \bar{\alpha}^0) - Q(h)\right)^2} \qquad \text{(A2.11)}$$

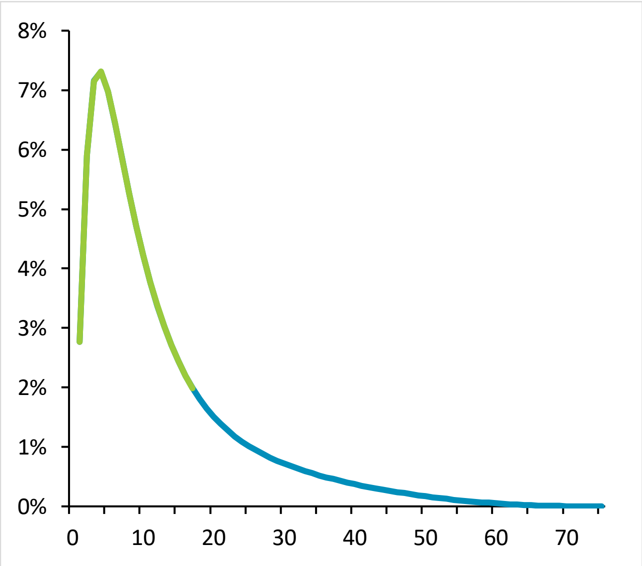
where $Q(h)$ denotes the empirical distribution. In figure A2.10 we show the mixed distribution based on the seed vector, $\bar{\alpha}^0$, against the hypothetical empirical distribution.

Figure A2.10: Initial fit of the mixed distribution based on seed values



Starting from the initial seed vector, $\bar{\alpha}^0$, we use solver in MS-Excel to search for the combination of the 6 parameter values that minimise the RMSE criterion. Convergence is achieved when the minimum RMSE is found. In our example the best fitting mixed distribution is shown in figure A2.11.

Figure A2.11: Best fitting mixed distribution

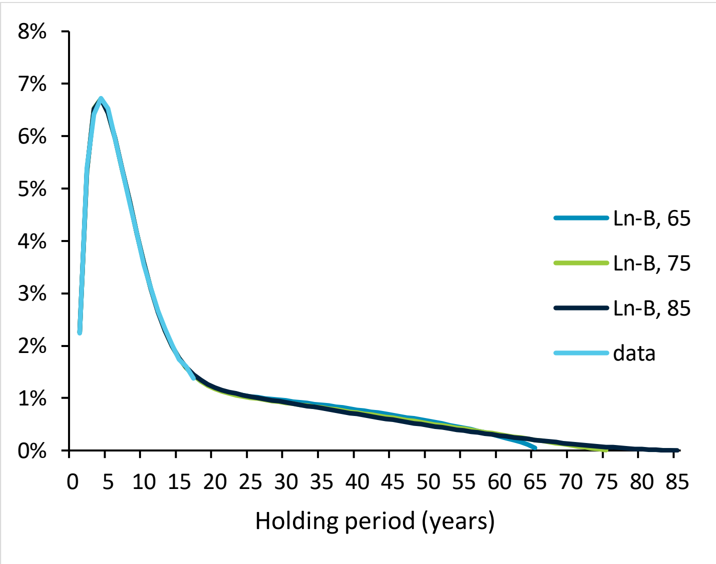


Appendix 3: Additional results

In this appendix we provide additional results from the distribution fitting exercise reported in Section 2. In Figures A3.1 – A3.3 we show fitted distributions based on alternative assumptions. The labels "*LN-B,##*","*B-B,##*" and "*W-B,##*" indicate, respectively, mixes of log-normal and beta distributions, mixes of two beta distributions and mixes of a Weibull and beta distributions. The number in the label indicates the maximum holding period assumed. Thus, the label "*B-B,85*" indicates the fitted distribution is based on a mix of two beta distributions with a maximum hold period assumed to be 85 years.

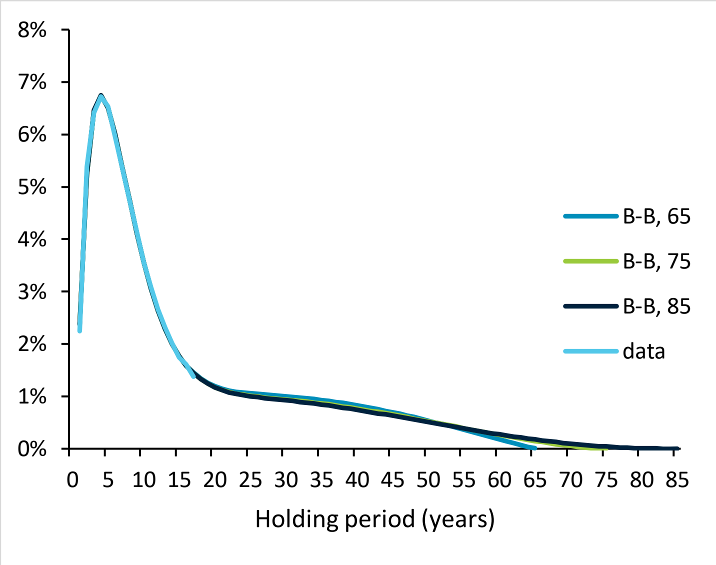
The beta-beta distributions generate results that are very similar to those for the log-normal-beta combination, although the RMSEs are higher. The Weibull-beta combinations produce lower average HPP across the combinations reported. However, the fit to the empirical distribution, indicated by the RMSE criterion, is not as good for the Weibull-beta combinations relative to the log-normal-beta and beta-beta combinations.

Figure A3.1: Mix of Log-normal and Beta distributions



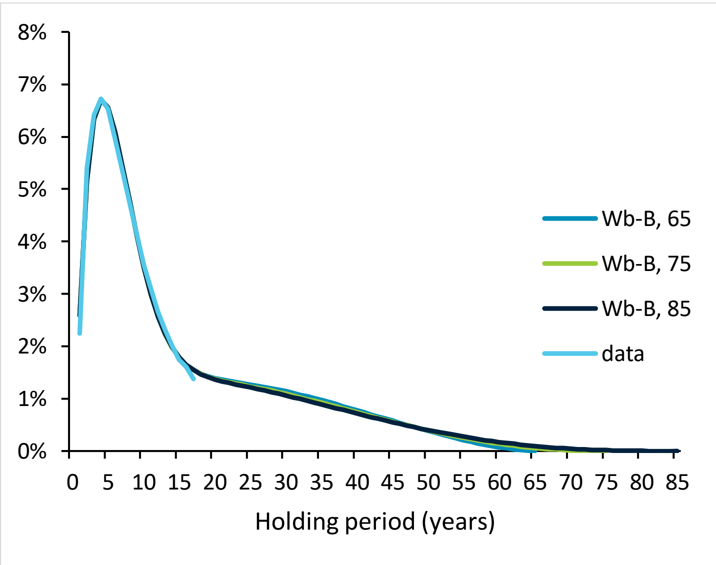
	Ln-B, 65	Ln-B, 75	Ln-B, 85
mean	16.8	17.2	17.1
median	9.7	9.7	9.7
RMSE	0.182%	0.180%	0.193%

Figure A3.2: Mix of Beta – Beta distributions



	B-B, 65	B-B, 75	B-B, 85
mean	16.5	17.0	17.1
median	9.7	9.7	9.7
RMSE	0.248%	0.239%	0.238%

Figure A3.3: Mix of Weibull – Beta distributions



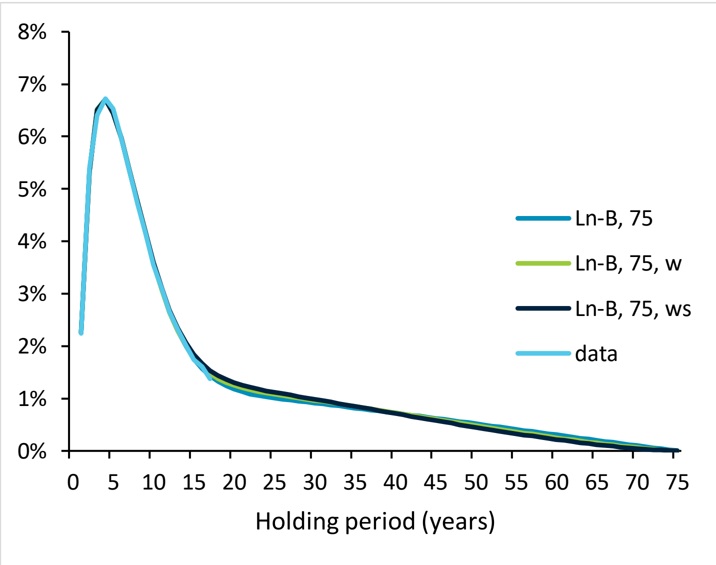
	Wb-B, 65	Wb-B, 75	Wb-B, 85
mean	15.7	16.0	16.2
median	9.6	9.6	9.6
RMSE	0.533%	0.524%	0.518%

As discussed in Section 2, the maximum holding period we have data for gets progressively shorter the more recent a vintage of transactions data is. For the vintage of properties that transacted in 2004-05, we have data on holding periods up to 17 years (for properties that transacted in 2020-21 for the first time since 2004-05). For the last vintage, which includes properties that transacted in 2010-11, the maximum holding period captured in the data is 11 years (for properties that transacted in 2020-21 for the first time since 2010-11).

In Figure A3.4 we show results based on different ways of dealing with the gradual lack of data for longer holding periods. The suffix "w" in the labels indicates that the RMSE is calculated with errors weighted to reflect the number of observations available to determine the empirical distribution.[8] The letter "s" indicates the shortening of the sample altogether, where for each vintage we drop the observations that capture the highest and second highest holding periods. Thus, for example, for the 2004-05 vintage we drop holding periods 16 and 17 (i.e., properties purchased in 2004-05 and sold in 2019-20 and 2020-21).

Weighting the errors or truncating the data does not change the shape of the distributions but results in a reduction in the average HPP. Reducing the weights on longer holding periods, the "w" approach, reduces the average HPP by 0.4 years. Focusing the fitting exercise on the shorter end of the distribution, the "ws" approach, lowers the mean HPP by an additional 0.3 years.

Figure A3.4: Truncated transaction data



	Ln-B, 75	Ln-beta, 75, w	Ln-beta, 75, ws
mean	17.2	16.8	16.5
median	9.7	9.7	9.7
RMSE	0.180%	0.163%	0.179%

[8] The empirical distribution that we are targeting is based on averaging across the 7 vintages the share of observations in each holding period. We have data for holding periods up to 17 years. If the errors associated with each holding period were equally weighted then each weight would be 1. For holding periods less than 12 years we have observations in each of the 7 vintages that can be used to estimate an average. For holding periods where there are fewer than 7 observations, the weight on the error is scaled down to reflect the fact that fewer vintages are available to estimate the average. Thus, the weights range from 1, where observations are available for all 7 vintages, to 1/7 for 17 year holding periods where only one observation is available.

Appendix 4: Home ownership model

This Appendix reproduces the home ownership framework set out in Warlters (2022).

Define:

$S_t = stock\ of\ dwellings\ at\ the\ beginning\ of\ year\ t$

$I_t = number\ of\ investor - owned\ dwellings\ at\ the\ beginning\ of\ year\ t$

$O_t = number\ of\ owner - occupied\ dwellings\ at\ the\ beginning\ of\ year\ t$

$T_t = number\ of\ transactions\ during\ year\ t$

$\gamma_t = \frac{I_t}{S_t} = investor\ share\ of\ the\ dwelling\ stock$

$1 - \gamma_t = \frac{O_t}{S_t} = owner - occupied\ share\ of\ the\ dwelling\ stock$

$\tau_t = \frac{T_t}{S_t} = transactions\ in\ year\ t\ as\ a\ share\ of\ the\ dwelling\ stock\ at\ the\ beginning\ of\ year\ t$

$\nu = growth\ rate\ of\ dwelling\ stock$

$\alpha = share\ of homes\ that\ are\ for\ sale\ that\ are\ purchased\ by\ investors$

$\beta_I = share\ of\ investor - owned\ homes\ that\ are\ sold\ each\ year$

$\beta_O = share\ of\ owner - occupied\ homes\ that\ are\ sold\ each\ year$

Assume that α, β_I, β_O and ν are parameters that do not vary unless policy is changed.[9] They can be thought of as long-run averages observed across the property market cycle.

These definitions yield:

$$S_{t+1} = S_t(1 + \nu)$$

$$T_t = \beta_I I_t + \beta_O O_t + \nu S_t$$

The investor-owned stock changes with each year's investor sales and purchases:

$$I_{t+1} = I_t - \beta_I I_t + \alpha(\beta_I I_t + \beta_O O_t + \nu S_t)$$

Dividing throughout by S_t we obtain:

$$(1 + \nu)\gamma_{t+1} = (1 - \beta_I)\gamma_t + \alpha[\beta_I \gamma_t + \beta_O(1 - \gamma_t) + \nu]$$

---

9 Warlters (2022) is concerned with how policy changes could influence the balance between owner-occupiers and investors in the dwelling stock (i.e., changes in γ_t). In that paper, tax policy could alter the shares of properties purchased and sold (i.e., α, β_I, β_O), while migration policy could influence the growth of the stock (i.e., ν). Policy changes to these variables would have an immediate effect on the determinants of annual transactions, but the stock would adjust gradually as it moved to a new equilibrium. Consequently, the model has time subscripts on the variables γ_t, and τ_t, but not on α, β_I, β_O and ν. In contrast, the current paper assumes no policy changes, and is only concerned with a housing market that is in equilibrium, so that the time subscripts may be dropped once the relationships between variables are established.

Defining $\Delta\gamma = \gamma_{t+1} - \gamma_t$, we observe that:

$$\Delta\gamma = 0 \; when \; \gamma_t = \gamma^*$$

$$\Delta\gamma > 0 \; when \; \gamma_t < \gamma^*$$

$$\Delta\gamma < 0 \; when \; \gamma_t > \gamma^*$$

where $\gamma^*$is a stable equilibrium investor share of the dwelling stock such that:

$$\gamma^* = \frac{\alpha(\beta_O+\nu)}{(1-\alpha)\beta_I+\alpha\beta_O+\nu}$$

Equilibrium transaction volumes as a share of the dwelling stock are the sum of sales by investors, owner-occupiers, and newly built dwellings:

$$\tau^* = \beta_I\gamma^* + \beta_O(1-\gamma^*) + \nu$$

We can simultaneously solve the last two equilibrium equations to find that:

$$\beta_I = \frac{\alpha\tau^*}{\gamma^*} - \nu$$

and

$$\beta_O = \frac{(1-\alpha)\tau^*}{(1-\gamma^*)} - \nu$$

Appendix 5: Conditioning the fitted distributions

This Appendix summarises the method we use to condition the mixed distributions fitted to Revenue NSW transaction data with information derived from home ownership model in Appendix 4. Our approach builds on the algorithm presented in Appendix 2, with two key differences:

- the distributions for owner-occupiers, investors and all buyers are fitted jointly, and
- the objective function includes arguments related to the parameters of the home ownership model.

Below we set out the details of this optimisation.

Optimization algorithm

Step 1: Select the maximum holding period, H.

Step 2: Select the family of theoretical distributions for the primary distribution, the secondary distribution, and the truncation function.

— The *Primary* distribution, $PR(h, \alpha_1, \alpha_2)$, is a 2-parameter distribution from either log-normal, beta, and Weibull.

— The *Secondary* distribution $SD(h, \alpha_3, \alpha_4)$, is a beta distribution.

— The *Truncation* function, $T(h, \alpha_5, \alpha_6)$, is the cumulative distribution function (CDF) of a beta distribution.

Step 3: Construct 3 mixed distributions:

— Candidate distribution of holding periods of owner-occupiers, F_{oo}:

$$F_{oo} = f\left(PR\left(h, \alpha_{oo,1}, \alpha_{oo,2}\right), SD\left(h, \alpha_{oo,3}, \alpha_{oo,4}\right), T\left(h, \alpha_{oo,5}, \alpha_{oo,6}\right)\right)$$

with $\bar{\alpha}_{oo} = \{\alpha_{oo,1}, \alpha_{oo,2}, \alpha_{oo,3}, \alpha_{oo,4}, \alpha_{oo,5}, \alpha_{oo,6}\}$ denoting the 6 parameters that define the primary and secondary distributions and the truncation function used to construct the mixed distribution for owner-occupiers.

— Candidate distribution of holding periods of investors, F_{inv}:

$$F_{inv} = f\left(PR\left(h, \alpha_{inv,1}, \alpha_{inv,2}\right), SD\left(h, \alpha_{inv,3}, \alpha_{oo,4}\right), T\left(h, \alpha_{inv,5}, \alpha_{inv,6}\right)\right)$$

with $\bar{\alpha}_{inv} = \{\alpha_{inv,1}, \alpha_{inv,2}, \alpha_{inv,3}, \alpha_{oo,4}, \alpha_{inv,5}, \alpha_{inv,6}\}$ denoting the 6 parameters that define the primary and secondary distributions and the truncation function used to construct the mixed distribution for investors.

— Candidate the mixed distribution to fit the data on holding periods across all buyers, F_{all}:

$$F_{all} = \alpha \times F_{inv}(h, \bar{\alpha}_{inv}) + (1 - \alpha) \times F_{oo}(h, \bar{\alpha}_{oo})$$

with $\bar{\alpha}_{all} = \{\bar{\alpha}_{inv}, \bar{\alpha}_{oo}\}$ denoting the 12 parameters defined above and α denoting the investor share of transactions discussed in Section 3 (as discussed in Section 3, $\alpha = 42.6\%$).

Step 4: Solve an optimisation problem.

— The choice variables are the 12 parameters that characterise the primary and secondary distributions and the truncation functions used to create the three mixed distributions, $\bar{\alpha}_{all} = \{\bar{\alpha}_{inv}, \bar{\alpha}_{oo}\}$.

— The objective function has two parts:

- The first part is the sum of the root mean squared error (RMSE) between the mixed distribution and the empirical distribution for owner-occupiers, investors and across all buyers. Let $Q_{xx}(h)$ denote the empirical distribution of holding periods for a specific group of buyers, with $xx = \{oo, inv, all\}$. The first part of the objective function is equal to:

$$\sum_{xx=\{oo,inv,all\}} RMSE_{xx} = RMSE_{oo} + RMSE_{inv} + RMSE_{all}$$

where:

$$RMSE_{xx} = \sqrt{\frac{1}{\bar{h}} \sum_{h=1}^{\bar{h}} \left(F_{xx} - Q_{xx}(h)\right)^2}$$

- The second part is the sum of the root squared error between the average holding periods associated with the three mixed distributions and the average holding periods calculated in Section 3:

$$\sqrt{(N_{oo} - \mu_{oo})^2} + \sqrt{(N_{inv} - \mu_{inv})^2} + \sqrt{(N_{all} - \mu)^2}$$

where $N_{xx}, xx = \{oo, inv, all\}$, is the average holding period associated with the mixed distribution, i.e.:

$$N_{xx} = \sum_{h=1}^{H} h \times F_{xx}$$

- The objective function is then equal to:

$$Objective = \sum_{xx=\{oo,inv,all\}} RMSE_{xx} + \sqrt{(N_{oo} - \mu_{oo})^2} + \sqrt{(N_{inv} - \mu_{inv})^2} + \sqrt{(N_{all} - \mu)^2}$$

— Use solver in MS-Excel to search for the values of the 12 parameters that minimize the objective function, i.e., that result is the smallest differences 1) between the candidate mixed distributions and the empirical distributions observed in Revenue NSW data and 2) between the average holding periods obtained from the candidate mixed distributions and the average holding periods obtained from the home ownership model.

Appendix 6: Preferred distributions

The following tables summarise our preferred estimates, derived in Section 3. The distributions provide the best fit with Revenue NSW transaction data assuming a maximum holding period of 75 years, a combination of log-normal and beta distributions, and incorporating mean values derived from aggregate property market data.

Central Moments for Holding Periods of Purchasers (HPP)

	All	Owner-occupiers	Investors
Mean	18.8	22.6	13.7
Median	9.7	10.5	8.8

HPP Distribution Functions

The x-th row on the left half of the following table indicates the share of purchasers expected to sell their dwelling in the x-th year of ownership. The x-th row on the right of the table indicates the share of purchasers expected to sell their dwelling in the x-th year of ownership or earlier.

Probability Distribution Functions				Cumulative Distribution Functions			
Year	All	Owner-occupiers	Investors	Year	All	Owner-occupiers	Investors
1	2.02%	1.10%	3.26%	1	2.02%	1.10%	3.26%
2	5.46%	4.40%	6.90%	2	7.48%	5.50%	10.17%
3	6.63%	6.12%	7.33%	3	14.12%	11.61%	17.49%
4	6.72%	6.55%	6.95%	4	20.84%	18.16%	24.44%
5	6.39%	6.37%	6.41%	5	27.23%	24.53%	30.85%
6	5.89%	5.91%	5.86%	6	33.11%	30.45%	36.71%
7	5.33%	5.35%	5.31%	7	38.45%	35.80%	42.02%
8	4.76%	4.77%	4.75%	8	43.21%	40.57%	46.77%
9	4.19%	4.21%	4.17%	9	47.40%	44.77%	50.94%
10	3.64%	3.68%	3.59%	10	51.04%	48.45%	54.53%
11	3.13%	3.20%	3.05%	11	54.18%	51.65%	57.58%
12	2.69%	2.77%	2.59%	12	56.87%	54.42%	60.16%
13	2.32%	2.38%	2.23%	13	59.18%	56.81%	62.39%
14	2.01%	2.04%	1.98%	14	61.20%	58.85%	64.37%
15	1.77%	1.74%	1.81%	15	62.97%	60.59%	66.18%
16	1.59%	1.48%	1.72%	16	64.56%	62.08%	67.90%
17	1.44%	1.26%	1.67%	17	66.00%	63.34%	69.58%
18	1.31%	1.06%	1.65%	18	67.31%	64.40%	71.23%
19	1.21%	0.90%	1.64%	19	68.52%	65.30%	72.87%
20	1.13%	0.75%	1.63%	20	69.65%	66.05%	74.50%
21	1.05%	0.63%	1.61%	21	70.70%	66.68%	76.11%
22	0.98%	0.53%	1.59%	22	71.68%	67.21%	77.70%
23	0.92%	0.45%	1.56%	23	72.60%	67.66%	79.26%

Probability Distribution Functions				Cumulative Distribution Functions			
Year	All	Owner-occupiers	Investors	Year	All	Owner-occupiers	Investors
24	0.86%	0.38%	1.52%	24	73.47%	68.04%	80.78%
25	0.81%	0.32%	1.48%	25	74.28%	68.36%	82.26%
26	0.77%	0.28%	1.42%	26	75.05%	68.64%	83.68%
27	0.72%	0.24%	1.37%	27	75.77%	68.89%	85.05%
28	0.68%	0.22%	1.31%	28	76.45%	69.11%	86.36%
29	0.65%	0.21%	1.24%	29	77.10%	69.31%	87.60%
30	0.61%	0.20%	1.18%	30	77.72%	69.51%	88.77%
31	0.58%	0.20%	1.11%	31	78.30%	69.70%	89.88%
32	0.56%	0.20%	1.04%	32	78.86%	69.91%	90.92%
33	0.53%	0.21%	0.97%	33	79.39%	70.12%	91.88%
34	0.51%	0.23%	0.90%	34	79.91%	70.35%	92.78%
35	0.50%	0.25%	0.83%	35	80.40%	70.60%	93.61%
36	0.48%	0.28%	0.76%	36	80.89%	70.88%	94.37%
37	0.47%	0.31%	0.70%	37	81.36%	71.19%	95.07%
38	0.47%	0.35%	0.63%	38	81.83%	71.54%	95.70%
39	0.47%	0.39%	0.57%	39	82.30%	71.93%	96.27%
40	0.47%	0.43%	0.51%	40	82.76%	72.36%	96.78%
41	0.47%	0.48%	0.46%	41	83.24%	72.84%	97.24%
42	0.48%	0.53%	0.41%	42	83.71%	73.37%	97.65%
43	0.49%	0.58%	0.36%	43	84.20%	73.95%	98.02%
44	0.50%	0.64%	0.32%	44	84.70%	74.59%	98.33%
45	0.52%	0.70%	0.28%	45	85.22%	75.28%	98.61%
46	0.53%	0.75%	0.24%	46	85.76%	76.04%	98.85%
47	0.55%	0.81%	0.21%	47	86.31%	76.85%	99.06%
48	0.57%	0.87%	0.18%	48	86.88%	77.72%	99.23%
49	0.60%	0.93%	0.15%	49	87.48%	78.64%	99.38%
50	0.62%	0.98%	0.13%	50	88.10%	79.63%	99.51%
51	0.64%	1.03%	0.10%	51	88.73%	80.66%	99.61%
52	0.66%	1.08%	0.09%	52	89.39%	81.74%	99.70%
53	0.67%	1.12%	0.07%	53	90.06%	82.86%	99.77%
54	0.69%	1.16%	0.06%	54	90.75%	84.02%	99.82%
55	0.70%	1.19%	0.04%	55	91.45%	85.21%	99.87%
56	0.71%	1.21%	0.04%	56	92.16%	86.42%	99.90%
57	0.71%	1.22%	0.03%	57	92.87%	87.64%	99.93%
58	0.71%	1.22%	0.02%	58	93.59%	88.86%	99.95%
59	0.70%	1.21%	0.02%	59	94.29%	90.07%	99.97%
60	0.69%	1.19%	0.01%	60	94.98%	91.26%	99.98%
61	0.67%	1.16%	0.01%	61	95.64%	92.42%	99.99%
62	0.64%	1.11%	0.01%	62	96.28%	93.53%	99.99%
63	0.61%	1.05%	0.00%	63	96.89%	94.58%	99.99%

Probability Distribution Functions				Cumulative Distribution Functions			
Year	All	Owner-occupiers	Investors	Year	All	Owner-occupiers	Investors
64	0.56%	0.98%	0.00%	64	97.45%	95.56%	100.00%
65	0.52%	0.90%	0.00%	65	97.97%	96.46%	100.00%
66	0.46%	0.81%	0.00%	66	98.43%	97.27%	100.00%
67	0.40%	0.70%	0.00%	67	98.84%	97.97%	100.00%
68	0.34%	0.60%	0.00%	68	99.18%	98.57%	100.00%
69	0.28%	0.48%	0.00%	69	99.46%	99.05%	100.00%
70	0.21%	0.37%	0.00%	70	99.67%	99.43%	100.00%
71	0.15%	0.27%	0.00%	71	99.82%	99.69%	100.00%
72	0.10%	0.17%	0.00%	72	99.92%	99.87%	100.00%
73	0.05%	0.09%	0.00%	73	99.98%	99.96%	100.00%
74	0.02%	0.03%	0.00%	74	100.00%	100.00%	100.00%
75	0.00%	0.00%	0.00%	75	100.00%	100.00%	100.00%

References

Corak, Miles and Andrew Heisz (1995). 'The duration of unemployment: a user guide'. Statistics Canada Working Paper 84.

Malakellis, Michael and Michael Warlters (2021). 'The economic costs of transfer duty: a literature review'. *Treasury Technical Research Paper 21-08*, NSW Treasury.

Warlters, Michael. (2022). 'Property tax reform and home ownership'. *Treasury Research Paper*, NSW Treasury, forthcoming.