Home equity insurance (preliminary version, please do not circulate)

Dag Einar Sommervoll, Gavin Wood

Abstract

We consider a home equity insurance where payouts are conditional on a house price index change and not by actual losses experienced by home owners. The efficiency of an index-based insurance depends on the covariance between individual transactions and market declines. We analyze insurance efficiency under various specifications of the underlying index and contract terms using a large data set of all market transactions of detached houses in the metropolitan area of Melbourne, Australia, for the time period 1990-2006. The payout efficiency measured by the percentage of payouts made to home owners experiencing a nominal housing transaction loss, tend to be robust under changes of the spatial aggregation of the house price index. In particular, a citywide index give a payout efficiency close to the average efficiency of a neighbourhood based index. The target efficiency, measured by the probability of a payout given a loss, is significantly higher for a neighbourhood based index.

JEL Classification: G13, R31, E32, R38
Keywords:

Affiliation:
Dag Einar Sommervoll is Research Fellow at Statistics Norway, Unit for Inequality, Taxation, and Consumer Behavior, Statistics Norway, Research Dept., P. O. Box 8131 Dep., N-0033 Oslo, Norway. E-mail address: dag.einar.sommervoll@ssb.no.

Gavin Wood is professor and director of Australian Housing and Urban Research Institute, RMIT, Melbourne Australia <more details>

1. Introduction

Owner occupied housing is wide spread in most OECD-countries and for many households the dwelling is their only financial asset. House purchases usually rely on mortgage financing, and homeowners use down payments on the principal as a way of saving. This pattern of saving is also encouraged in some countries by a lenient taxation of owner occupied housing. At the same time do investments in housing markets come with considerable risk as house prices tend to vary considerably over time.

In contrast to most other financial risks, few vehicles for hedging, trading risk, or risk diversification exist for home owners, and none are wide used. In asset markets the risk associated with future price movements are routinely traded in derivative markets. The risk of a purchase of an asset today can be hedged by taking a short position in the futures market of the same asset.

The basic requirement for a derivative market of futures, is agents that are willing to long and short positions. The housing market meets this requirement. Home owners may hedge risk by taking a short position in a market of housing futures. Real Estate developers and future home owners may hedge their risks by taking a long position. So far a few markets for housing derivatives has emerged <(Swidler, Basel)>. One reason for this may be that such markets and institutions take time to evolve, and the limiting scope we see today may be just the humble beginning of highly liquid derivative market which in the future will not only serve as a way to mitigate risk for homeowners, but provide opportunities to hedge risks from other parts of the economy. One potential obstacle for such products may be the economic literacy of households, and some may feel reluctant to take a financial position in a derivative market which they do not fully understand. Furthermore, a hedge comes with a price. Some of the potential housing return is traded for less severe losses in case of market decline. To what extent a particular homeowner is likely to forsake future housing returns in order to be on the safe side depends on their attitude towards risk. The recent global financial crisis and falling house prices may make households in general more aware of the potential downside of unhedged risk.

The housing investment risk may be reduced in other ways than hedging in derivative markets. Most home owners purchase a variety of insurances. In particular, most home owners routinely insure their home, both the house and contents against unfortunate events like fire or "break ins". A home equity insurance may formulated along similar veins giving payouts in the case of house price down turns. One obvious challenge is that a future selling price of a house is not a random event like fire or burglary. Both maintenance and effort to find the right buyer is known to affect the selling price (Genesove and Mayer 1993). One way to deal with this moral hazard[1] problem is to make payouts conditional on house market movements, that is house price index changes, and not individual market transactions. This is our point of departure.

We consider home equity protection by a purchase of a home equity insurance. and ask:
How efficient is an index based insurance in targeting incurred losses?
To what extent does the efficiency depend on the index construction and rules regarding legitimate claims?

We address these questions by considering the performance of such an insurance for given the real housing market transactions of detached houses in the metropolitan area of Melbourne 1990-2006. We consider the number of claims, the size of payouts if an insurance was in place. In particular, we consider various specifications of the index, and various rules regarding legitimate claims.

Our computations are based on historic transactions in a real housing market, and do not capture the effect of an insurance on reservation and transaction prices. Potential insurance payouts are likely to affect behavior. In this sense our results may be viewed as static and "ideal" since no adverse selection or moral hazard related strategic considerations affect market outcomes.

Both adverse selection and strategic selling times can hurt the profitability (in case of private insurers) or challenge sponsorship (in the case of governmental initiative). Though these supply side challenges are important for the implementation and success of an insurance product of this type, we consider only the demand side in this paper, and limit the analysis to what extent an index based home equity insurance, manage to provide home equity protection.

We find that the payout efficiency, that is to what extent payouts are made to home owners that incurred losses, tend to be robust under changes of the spatial aggregation of the house

[1] It may be argued that this is not moral hazard in a strict sense.

price index. In particular, a citywide index give a payout efficiency close to the median efficiency of a neighbourhood based index (64,3%). The target efficiency on the other hand, the likelihood of getting a payout when experienced a loss, is much higher for low spatial aggregation. The temporal aggregation of the index within plausible limits, say ranging from monthly to half yearly indices does have little effect on payout efficiency, but indices defined on neighborhoods the payout efficiency can at times be quite sensitive to small changes in temporal aggregation.

Home equity insurance where payouts is conditional on a house price index movements and not the house price movements of the homeowners house, can also be viewed as a hedge. A home equity insurance that covered all losses of individual home transaction could in the same wording be viewed as a perfect hedge. In this sense our approach is similar to (Bertus, Hollans et al. 2008), though we do not share the investor perspective. We consider home equity insurance as a way for households to reduce risk without taking in the larger perspective of risk sharing and portfolio diversification.
< Also the Syracuse working paper here>

The paper is organized as follows. Section 2 starts with a brief discussion of existing home equity insurance programs, and proceeds to discuss measures of efficiency. We propose a new efficiency measure, target efficiency, to supplement the existing (payout) efficiency measure previously used in the literature (Caplin, Goetzmann et al. 2003).
Section 3 describes the data set of all detached house sales in the metropolitan area of Melbourne 1990 to 2006. In section 4 we consider home equity insurance efficiency under different aggregation over space and time, and continue to the efficiency under various specification of maturity times. Section 6 concludes.

2. Home equity insurance

2.1 Historic back ground

The first US insurance project was in 1978 Oak Park Illinois, and Oak Park residents could ensure the home equity against future loss (80 percent coverage) for a one time fee of 175 dollars. The intension of the program was to prevent urban decline, and as such it proved very successful. Prices did not fall, and no insurance claims were filed, during the entire length of the program. The number of insurance holders remained low, and was never higher than 151 in all. The Oak park program is best described as small and local. Moreover, potential payouts should be financed by a general property tax. In other words, the insurance was not standard commercial product where risk was traded. It was closer to a common good, where every resident had to contribute.

Interestingly, the Oak Park Insurance Program (OPIP) was intended to cover only local fluctuations. In particular, a house price index fall in excess of 5 percent in the Mid West region of Illinois, would imply a temporal suspension of the program. A potential claim was based on actual transaction prices, thus susceptible to moral hazard problems. To prevent this, elaborate rules applied for selling a house. In particular, there was a chance of write down of the guaranteed value in case of poor maintenance. The complexity of a potential claim may have attributed to the success of the program. It deterred insurance holders from making claims, but provided security through the option to file a claim. Most likely, the limited scope of the program, insurance against only local prices down turns may have made the program less appealing to home owners in general.

In 1998 housing economists from Yale and NYU was invited to tailor a home equity insurance program for Syracuse(Caplin, Goetzmann et al. 2003). The city of Syracuse had declining population and falling house prices through out the 1990s. Part of the problem was a “bank run” type of a problem, sell your house before prices fall even more. A home equity insurance could, if properly designed, give Syracuse residents security from future losses. And potentially reverse the trend, if the larger part of the problem was lack of confidence in the housing market.
The Syracuse insurance program was inspired by the OPIP, but differs on several important points. To address the moral hazard problem, losses are based on changes in local prices indices instead of actual transaction. If you purchase a home for 100 000 dollars, and the price index falls by 10 percent. Selling the house at entitles you to an insurance pay of to 10 000 dollars, irrespective of your selling price. The financing of the insurance, is a fund of 5 million USD provided by the state of New York in addition to a one time fee (1.5 percent) of purchase price.

2.2 Definition of an index based insurance scheme
In principle any insurance product where payouts are conditional on movements of an index rather than individual price movements is an index based insurance scheme. We limit the discussing where payouts are proportional to the index change in such away that, in the event of identical price movements a given insured house and the index, the home owner is fully covered by the insurance.

In other words the insurance payout $\pi(i)$ for a given house i is given by:

$$\pi(i)=[I(t_0(i))-I(t_1(i))]\ p_0(i)/\ I(t_0(i)),$$

where $p_0(i)$ is the purchase price at time $t_0(i)$, and $I(t_0(i))$ refer to the housing price index at time $t_0(i)$, and $I(t_1(i))$ to the index at time $t_1(i)$,

Consider the following numerical example. You buy a house i at time $t_0(i)$ for a price, $p_0(i)$ = 100 000 USD, and a later time $t_1(i)$ you sell your have for 90 000. That is you sell your house for 10 percent less (10 000). At the same time the housing market index is down 8 percent. If you are home equity insured, you will recover 8 percent of your losses or 8 000 USD.

An index based insurance may or may not be and efficient way to mitigate housing market risk. In the following we will briefly outline various measures of how efficient an index based insurance is in reducing risk.
Payout efficiency is defined to be the ratio between the size of payouts to people who suffered losses on the housing market and the size of all payout:

$$E_{payout}=\Sigma_{loss}\ \pi(i)/\ \Sigma_{all}\ \pi(i)$$

The average loss coverage (given a payout) is defined to be:

$C_{payout} = \Sigma_{loss}\ \pi(i)/\ \Sigma_{loss|\ payout}\ [p_0(i) - p_1(i)]$

In the following analysis and discussion we will largely rely on these measures.[2]

The E_{payout} captures to what payouts end up in the pockets of house sellers that experienced losses, but E_{payout} nor the other measures captures to what extent a loss result in a payout.

We propose the following (target) efficiency measure:

$E_{target}=\Sigma_{loss\ and\ payout}\ i/\ \Sigma_{loss}\ i$

And interpret E_{target} as an estimate of the probability of payout given a loss.

3. Data description

The data set consists of all housing market transactions of detached houses in the metropolitan area of Melbourne, Australia in the time period 1st of January 1990 until 31 of December 2006. From a raw data set of 970 502 sales, 505 252 sales were of houses sold more than once and less than 9 times,[3] giving 210 118 houses. From this data set a net sample of 176 861 houses was extracted, giving in total 223 461 transaction pairs (see appendix for details). All transactions in the net sample have a number of attributes, including GPS coordinates, post code, and neighbourhood.

Nominal losses occurred for 12,5 percent (27 959 of 223 461) of the transaction pairs. Table 3.1 gives the distribution of transaction pairs in the time period 1990-2006. This overview show that sales involving nominal losses have occurred through out the time period 1990-2006.

Table 3.1. Distribution of transaction pairs for the time period 1990-2006. Rows first sale. Columns second sale. Second row in each year gives the number of sale pairs with nominal loss. Boldface numbers with minus sign indicate if the difference of corresponding pair of Case-Shiller indices[a] was negative.

[2] In Caplin, A., W. Goetzmann, et al. (2003). Home Equity Insurance: A pilot Project. Yale ICF Working Paper. Two more measures are discussed: the payout ratio, $P_{ratio}=\Sigma_{all}\ \pi(i)/\ \Sigma_{all}\ p_0(i)$, and the loss ratio $L_{ratio} = \Sigma_{loss}\ [p_0(i) - p_1(i)]\ /\Sigma_{all}\ p_0(i)$. These two measures are related to the ones discussed above in the following way: $E_{payout}\ P_{ratio} = L_{ratio}\ C_{payout}$

[3] Transactions of houses sold ten times or more excluded from the analysis. For further details see Section A.2 of the

	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
1990	273	989	1454	1553	1644	1530	1639	1866	1498	1491	1203	1066	933	749	637	620	383
	-123	***-597***	***-925***	***-948***	***-833***	***-906***	***-943***	***-789***	435	254	80	33	8	9	5	5	7
1991		205	890	1377	1684	1552	1666	1981	1677	1543	1329	1149	1011	813	696	682	445
		-78	***-393***	***-649***	596	736	786	594	301	116	38	14	10	4	5	7	10
1992			760	916	1609	1578	2643	2158	1923	1825	1589	1402	1130	980	857	776	513
			-250	***-334***	531	715	1208	574	283	117	46	27	23	11	4	12	3
1993				799	934	1354	1669	2182	1986	1900	1742	1478	1291	1106	896	848	528
				-62	269	570	667	521	231	84	33	22	12	4	6	8	1
1994					299	917	1613	2237	2123	2187	2073	1751	1521	1204	1022	1003	594
					-79	***-440***	***-731***	581	255	108	34	24	12	7	8	10	4
1995						305	988	1776	1790	1966	1875	1677	1508	1224	1027	1052	659
						-83	***-372***	340	154	98	54	24	16	7	7	10	3
1996							963	1159	1754	2004	2064	1906	1691	1417	1246	1250	743
							237	189	176	95	63	29	15	11	15	12	7
1997								371	1348	2265	2557	2497	2258	1831	1655	1645	1078
								69	202	125	78	36	34	22	19	18	9
1998									399	1387	2175	2455	2149	1912	1647	1641	1072
									103	145	113	63	36	20	21	24	8
1999										324	1596	2560	2442	2109	1948	1995	1308
										59	189	98	59	42	41	30	17
2000											416	1648	2399	2355	2210	2344	1543
											65	136	77	65	65	46	24
2001												328	1642	2763	2666	2817	1925
												55	113	115	117	111	51
2002													348	2106	2966	3254	2259
													58	182	209	198	100
2003														443	1775	2923	2237
														87	333	484	312
2004															395	1522	1807
															-89	336	333
2005																357	982
																88	217
2006																	219
																	-38

[a]Case-Shiller indices to be disussed in more detail in Section 4 and in the Appendix

Table 3.2 shed more light more light on the price movements in general. We see a drop in median prices in the early 1990ies, a recovery in 1996-1997 and considerable appreciation for the rest of the time period[4]. The increase of average holding time (time between sales) over the time period 1990-2006 is largely due to the data gathering process. Holding times for the year 1990 is maximum one year. However, the difference in holding times for sales with loss in 2006 (2.8), in contrast to the average holding time for pairs with nominal gain (6.4) is reflects market movements and are important when considering home equity insurance based on index movements.

Table 3.2. Summary statistics by year.

Year	Number of sales	Number of resales	Number of resales with nominal loss	Percent resales with loss	Median price	Average holding time	Average holding time if gain	Average holding time if loss
1990	19 676	273	74	27	129 000	0.4	0.3	0.5
1991	19 321	1 194	438	37	125 000	1.0	0.9	1.1
1992	22 451	3 104	1 100	35	122 500	1.4	1.1	1.6
1993	21 169	4 645	1 294	28	122 500	1.9	1.5	2.3
1994	21 970	6 170	1 572	25	126 000	2.6	2.4	2.8
1995	20 049	7 236	2 344	32	125 000	3.1	2.9	3.3
1996	23 652	11 181	3 707	33	125 000	3.4	3.2	3.7

[4] Note that yearly median prices is a quite crude measure of temporal price variation. A more refined index like the Case-Shiller index used on lower aggregation over both space and time, add substantially to understanding of the price movements for the time period in question.

1997	26 352	13 730	2 834	21	139 000	4.1	4.0	4.5
1998	24 636	14 498	2 001	14	150 000	4.4	4.4	4.9
1999	26 291	16 892	1 412	8	166 500	4.8	4.8	5.1
2000	26 788	18 619	1 203	6	183 000	5.0	5.0	3.9
2001	27 512	19 917	1 019	5	220 000	5.2	5.3	3.7
2002	27 184	20 323	904	4	250 000	5.5	5.6	3.4
2003	25 565	21 012	971	5	280 000	5.5	5.6	2.6
2004	23 812	21 643	1 191	6	300 000	5.7	5.9	2.5
2005	25 460	24 729	1 534	6	310 000	6.0	6.2	2.8
2006	18 434	18 295	1 204	7	320 000	6.2	6.4	2.8

:

4. Analysis

In the previous section we gave a brief overview of transaction pairs and median price movements for the time period in question. In this section we will consider the efficiency of a home equity insurance based on a Case-Shiller index[5]. In particular, we will discuss temporal and spatial aggregation of the index, since the correlation between the individual house losses and index decline is essential for an insurance product of this kind. Moreover, we will consider maturity times, the minimal holding time in order to file an insurance claim. In the light of Table 3.2, maturity times is potentially important since losses in the 2000s were associated short holding times.

The Case-Shiller index and efficiency

A Case-Shiller index tend to be data intensive as only transaction pairs is considered. Moreover, the index is prone to small sample biases (Meese and Wallace 1997), (Sommervoll 2006), (Sommervoll and Wood)), and at worst may give misleading and highly volatile estimates of price movements. From a home equity insurance perspective this contingency is especially important to avoid, since consumer confidence in the insurance product is essential. At the same time, considerable spatial price variation, and henceforth (sub)index variation are expected in most housing markets. In particular, an aggregation of over neighborhoods which are strikingly different with respect to housing characteristics, and are potentially expected to experience separate price trajectories, may make the insurance product unattractive to some home owners, and be seen as a arbitrage opportunity for others.

We address this problem by considering two extremes. An insurance scheme based on city wide, or a neighborhood quarterly index Case-Shiller index. The neighborhood index is based on a division of the Metropolitan area of Melbourne into 108 neighborhoods. The division is done in such a way that no neighborhood has to few repeated sales (see Appendix for details, xxxcheck appendix), and at the same time reflect neighborhood borders that are adhered by the agents in the housing market.

[5] For details on the Case-Shiller index see the Appendix

Figure 4.1. Case-Shiller indices. City wide and four selected neighborhoods

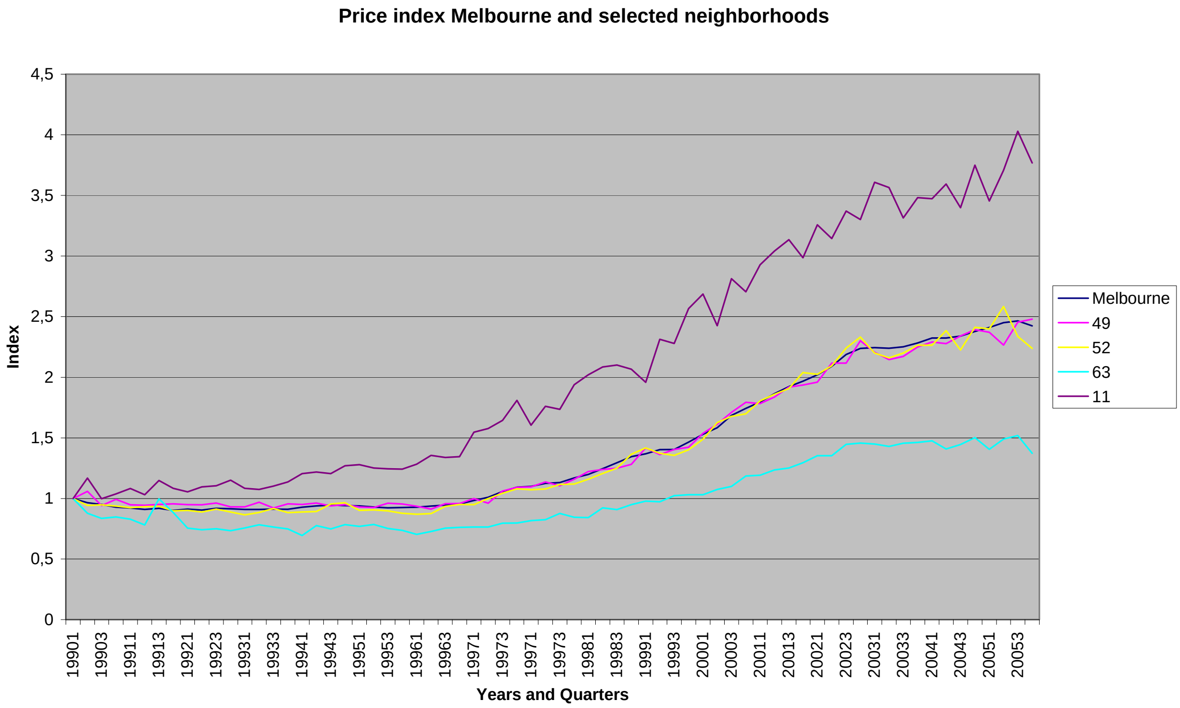

Figure 4.1 displays the city wide index as well as a four selected neighborhood. (Neighborhood, 49, 52, 11, and 63) We see that neighborhoods like 49 and 52, move closely to that of the city wide index. This is in contrast to the neighborhood 11 and 63. Neighborhood 11did not experience recovered the early 90ies down turn and houses appreciated much faster than the market as a whole. Needless to say, and insurance based on the city wide index would have proved attractive for home owners in this neighborhood. In neighborhood 61 home owners continued to sell in submarket which did not recover before the year 2000. The Melbourne market recovered three years earlier. An insurance based on a city wide housing index, would not have covered any of the losses experienced during these three years.

These broad considerations does not share light on how much efficiency is compromised by relating payouts to the city wide index in contrast to a much more spatially disaggregated index. Table 4.2 displays a comparison of the two scenarios. We see that the payout efficiency is comparable for the two (63.6 percent for city wide in contrast to the neighborhood median of 64.3). This implies that the correlation between individual price movements and the corresponding price movements do not get considerably higher (0.56 versus median 0.78), by using a spatially more disaggregated index.

However, there is a significant difference between the target efficiency of the two scenarios. The number of payouts in the neighborhood case is much higher, and interpreted target efficiency as a probability, the probability if payout given a loss is much higher in the neighborhood case (45.5 percent in contrast to 24.3 percent).

Table 4.2. Payouts and efficiency. Insurance based on a Melbourne index or Melbourne neighborhood indices.

Spatial division	No. Payouts	No. Payouts loss	No. Payouts win	Av. Loss Coverage[1]	Payout Efficiency[1]	Target efficiency[1]
Melbourne	12 686	5687	5 879	28.8	63.6	24.3
Neighborhoods	22 170	12 691	9 479	43.7	64.3	45.5

In principle could the low payout efficiency for both the city wide and the neighborhood index be due to high temporal aggregation. If true, letting the payouts be conditional on a bimonthly index, rather than a quarterly improve efficiency. This is not the case. As a sensitivity check of temporal aggregations all possible possible equidistant partitions of the time interval in question from 34 time periods (half year indices) to 136 (half quarter indices) did show no upward trend. Moreover the mean payout efficiency and mean target efficiency is close to the quarterly one (63.9 (0.04) and 22.3 (0.05) respectively). For the neighbourhood indices sample sizes are generally to small to allow for a considerable lower temporal aggregation. For some neighorhood efficiency tends to be unstable in the sense that a change of the number of time periods from 68 to 60, may off set an efficiency change of several percentage points. This unstability was primarily not due to profound changes in the underlying index and then a symptom of undersmoothing of the estimated index (Sommervoll, Sommervoll&Wood). It was a largely driven by a small sample of legitimate claims, and a small change in the number legitimate claims may have a considerable impact on efficiency either up or down depending on the underlying sales where nominal losses or not.

Rules for insurance claims and efficiency

The attractiveness of an insurance depends on the rules regarding legitimate claims. As discussed in Section 2, the Oak park insurance program had rather elaborate rules limiting the prospects of payouts when incurred losses in the Oak park housing market. The insurance schemes discussed so far in this paper resembles an American Style futures contract, where the holder of the contract can make a claim in any point in time. In this section we will consider insurance contracts that are closer to a European style futures contract. That is the insurance contract has a maturity date, and no claims can be made prior to this date. Of prime interest is whether longer holding times increase the efficiency at least in times of general appreciation, since shorter holding times are expected to carry more idiosyncratic transaction noise relative to movements in the index. We consider holding times at least 2, 3, 4, 5 years.

Table 5.1 Efficiency and maturity dates for insurance claims. Insurance based on the Melbourne citywide index.

Maturity in Years	No. Payouts	No. Payouts loss	No. Payouts win	Av. Loss Cov.	Payout Efficiency	Target efficiency	Conditional Target Efficiency
1	10002	5687	4315	28.3	64.6	20.3	23.8
2	6465	3833	2632	28.6	64.2	13.7	20.8

3	4572	2715	1857	26.5	63.1	9.7	20.4
4	3329	1968	1361	22.3	61.9	7.0	21.5
5	1950	1196	754	22.8	63.3	4.3	20.5
6	637	390	247	18.6	58.7	1.4	11.4

Table 5.1 Efficiency and maturity dates for insurance claims. Insurance based on the Melbourne neighborhood indices. Median average loss coverage if payout. Median Payout and Target efficiency. Median Conditional Target efficiency[a].

Maturity in Years	No. Payouts	No. Payouts loss	No. Payouts win	Av. Loss Cov.	Payout Efficiency	Target efficiency	Conditional Target Efficiency
1	18 041	11 116	6 875	41.9	68.8	38.4	46.6
2	13 655	8 823	4 832	41.4	69.9	28.8	47.8
3	9 632	6 471	3 161	42.0	73.1	21.7	49.1
4	6 419	4 393	2 026	42.5	76.0	16.9	52.0
5	3 812	2 646	1 166	49.9	75.8	9.7	49.6
6	1 974	1 352	622	52.3	76.9	4.9	44.8

[a]Conditional target efficiency is defined to be the target efficiency in the subset of transaction pairs with holding times exceeding the maturity time.

5. Conclusion

Many households have most of their equity in the housing market. As housing markets tend to be volatile, self owned housing comes with considerable risk. A market down turn may heart individual households, and at times have serious ramifications for other parts of the economy, as loss of home equity can affect consumer spending. In this paper we considered index based home equity insurance as a way to reduce housing investment risk for home owners.

The key question for any insurance product to what extend it mitigates the risk of the insurance buyer. In the case of the index based home equity insurance this translates to the correlation between individual house price movements and the corresponding index movements.

References

Bailey, M. J., R. F. Muth, et al. (1963). "A Regression Method for Real Estate Price Index Construction." Journal of the American Statistical Association **58**(304): 933-942.

Bertus, M., H. Hollans, et al. (2008). "Hedging House Price Risk with CME Futures Contracts: The Case of Las Vegas Residential Real Estate." Journal of Real Estate Finance and Economics(37): 265-279.

Caplin, A., W. Goetzmann, et al. (2003). Home Equity Insurance: A pilot Project. Yale ICF Working Paper.

Case, K. E. and R. Shiller (1989). "The Efficiency of the Market of Single -Family Homes." The American Economic Review **79**(1): 125-137.

Genesove, D. and C. Mayer (1993). Understanding the Price-Volume Correlation: The Impact of the Real Estate Cycle on the Decision to Sell. M.I.T. Working Paper. Boston, M.I.T.

Meese, R. and N. Wallace (1997). "The Construction of Residential Housing Price Indices: A Comparison of Repeated Sales, Hedonic Regression, and Hybrid Approaches " Journal of Real Estate Finance and Economics **14**(1-2): 51-74.

Sommervoll, D. E. (2006). "Temporal Aggregation in Repeated Sales Models." Journal of Real Estate Finance and Economics **33** (2): 151-165.

Sommervoll, D. E. and G. Wood Spatial and Temporal Aggregation in Repeated Sales Models.

Appendix

A.1 Repeated sales methodology

The classical repeated sales model of Bailey, Muth, and Nourse (the BMN-model)(Bailey, Muth et al. 1963) is given by regressing the difference in log sale prices for same homes on a set of time dummies, as presented in equation (1).

$$(1) \qquad \log(p_{it}) - \log(p_{is}) = \gamma_2 D_{i2} + \gamma_3 D_{i3} + ... + \gamma_m D_{im} + \varepsilon_{it}, \qquad i \in I; t, s, \in \{2,...,m\}, D_{it} \in \{-1,0,1\}$$

where p represents sale price, D is a dummy variable indicating first sale, second sale or no sale, t is the time period in which the second sale was undertaken, s the time period in which the first sale was undertaken (and thus $s<t$), subscripts i refer to the sale of a given object in the set of all repeated sales I such that i refers to an object sold exactly two times, γ's are index parameters to be estimated, and ε is an error term with zero-mean, and constant variance. The dummy variable D is set to +1 in the second period it was sold and –1 in the first period it was sold for each object, unless this is the first time period, where the dummy variable is set to 0. The later corresponds to a normalization of prices to 1 in the first period. The time dummies correspond to a partition of a time interval [0,T] into m parts, where all the transactions occur.

In the case that the error terms are independently normally distributed with zero mean, the least square estimates of (1), give minimum variance and (linear) unbiased estimates of the γ's. However, if the error terms increase over time, this is no longer true. (Case and Shiller 1989) argued that error terms are likely to be higher for dwellings where the time interval between sales is larger. They employed a variant (WRS-index) in which the time dependence of the variances is estimated and the model is estimated a second time using weighted least squares to correct for heteroskedasticity. For the simulation analysis in this article only the BMN-index is considered.

Under the assumption that the model is correct, there are several important facts that follow immediately. First the BMN-estimators of the log price coefficients are normally distributed, unbiased stochastic variables, irrespective of sample size. Thus, in this setting seriously biased estimates are apriori rare. However, with price movements within each time period corresponding to a time dummy, as is the case in true housing markets and in the simulation analysis below, the constancy condition within each time period is violated. Asymptotically, this is of no concern, since the BMN-estimator is ordinary least squares (OLS) and asymptotic normality follows from the i.i.d. hypothesis of the error terms. In a finite and small sample scenario, it is unclear to what extend asymptotic properties prevail.

The second theoretical point is that there is no notion of time in the model. So no specific partition of the time interval [0,T], is favored from the model itself. Since sparseness of a data set depends heavily on the chosen temporal aggregation, a data set tends not to be a priori sparse, but merely sparse relative to a given temporal aggregation. In the simulation analysis this may be utilized to isolate effects that are merely driven by temporal aggregation, by estimating the model on the same data set varying the aggregation only.

A.2 The Data set

Table A.1 Preparation of the data set. Details.

Data operation	Number of sales
All transactions	970 502
Price between 20 000 and 500 000 USD	968 848
Number of sales, where the house has been sold between 2 and 9 times	505 252
Number of houses sold between 2 and 9 times	210 118
Number of houses with distinct sale times	204 764
With geocodes	204 702
Random pairs with less than 30 percent yearly appreciation	176 891

Table A.2 Number of transactions in the net sample of all multiple transactions.

No. of transactions	2	3	4	5	6	7	8	9
No. of houses	128517	31763	5990	872	101	14	2	0

Figure A.3 Spatial division of the Melbourne metropolitan area.
<To be included >