A NEW HEDONIC METHOD FOR MEASURING MORTGAGE DEFAULT RISK

KAI LIN[+], DR. AINSLIE YUEN[+] AND CHRISTOPHER JOYE[+]

[+]Coolabah Capital Investments

Abstract

Measuring and monitoring the true level of mortgage delinquencies across an economy is essential for asset pricing and financial system stability. Yet public measures of mortgage default risk almost always use simple averages across pools of individual assets, including balance-sheet loans or indices tracking default risk across portfolios of residential mortgage-backed securities (RMBS). These approaches are, like median house price indices, afflicted by compositional biases that can lead to spurious inferences regarding the direction of default rates. Sources of bias include artificial changes in default rates attributable to: increases in the volume of new loans being written or securitised RMBS added to indices; changes in the proportion of transactions with higher loan-to-value ratios (LVRs); the introduction of less seasoned RMBS transactions with a lower weighted-average loan age; and/or borrower characteristics that have higher probabilities of default (e.g. tilts towards investment borrowers). To address this problem, we have developed the first known hedonic regression-based indices of mortgage default risk that explicitly control for compositional biases through the models' characteristic-based independent variables. Whereas simple average measures of default rates across securitised loan portfolios have declined in recent years, which suggests that the risk of loss has been declining, our hedonic mortgage default index implies exactly the opposite: that is, compositionally-adjusted default rates have, in fact, been increasing sharply in recent times.

A New Hedonic Method for Measuring Mortgage Default Risk

Kai Lin[+], Dr. Ainslie Yuen[+] and Christopher Joye[+]

[+]Coolabah Capital Investments[1]

Being able to measure and monitor the true level of mortgage delinquencies across an economy is essential for investors and regulators interested in asset pricing and financial system stability, amongst other things. Prior to the 2008 global financial crisis, rising mortgage default rates in the United States were an important leading indicator of subsequent losses. Changes in arrears rates are also vital for the valuation of asset-backed securities sold via securitisation.

And yet public measures of mortgage default risk almost always use simple averages across pools of individual assets. In Australia, one such example is the Standard & Poor's Global Ratings Mortgage Performance Index, known as the SPIN Index.[2] These weighted-average or median measures of default rates are, like median house price indices, afflicted by severe compositional biases, which can lead to spurious inferences regarding the direction of default rates.

New residential mortgage-backed securities (RMBS) issues that are, for instance, added to the SPIN Index sample typically have no defaults in them. This is because a new RMBS deal will typically start with only performing home loans. That means that a significant increase in RMBS issuance, such as that which has occurred in Australia in recent years, can artificially reduce the SPIN Index's reported arrears rate at a time when the true or underlying default rate is actually climbing.

There are other sources of potential bias with simple weighted-average or median default indices. Even if RMBS issuance patterns are constant over time, a change in the composition of the deals, such as a jump in the proportion of transactions with higher loan-to-value ratios (LVRs); less seasoned loan pools (i.e. portfolios with a lower weighted-average loan age); and/or borrower characteristics that have higher probabilities of default (e.g. tilts towards investment borrowers, interest-only borrowers, and/or sub-prime borrowers) can result in a simple weighted-average or median mortgage default index producing misleading results.

To address this problem, we have developed the first known hedonic regression-based indices of mortgage default risk that explicitly control for compositional biases through the models' characteristic-based independent variables, which include the time since the RMBS

[1] We are appreciative of the feedback of the Coolabah Capital Investments' executive team.

[2] RMBS Arrears Statistics S&P Global Ratings.

transaction was consummated, the weighted-average age of the loans, and the weighted-average LVR, amongst other factors.

Whereas the SPIN Index suggests that default rates across Australian RMBS transactions have declined in recent years, which has been used in many issuers' marketing materials to promote the security and safety of these deals, our hedonic mortgage default index implies exactly the opposite: that is, compositionally-adjusted default rates on Australian RMBS deals have, in fact, been increasing sharply in recent times (see Figure 1).

While this finding conflicts with the claims of credit rating agencies that monitor RMBS default rates, it reconciles closely with the arrears rates reported by the Reserve Bank of Australia (RBA), which tracks the performance of all bank balance-sheet loans (as opposed to just the sample of RMBS issues), and the higher impairments disclosed by Australia's biggest banks.[3] Indeed, our hedonic mortgage default index correlates closely with the RBA's published mortgage default series, which makes sense given that the RBA data is not impacted by biases induced by the frequency of RMBS issuance (see Figure 1).

Another finding is the presence of strong seasonality in mortgage default rates, which we quantitatively address via the application of seasonal-adjustment methodology. To the best of our knowledge, the SPIN Index is not currently seasonally-adjusted.

Going forward, we would encourage researchers to develop and use hedonic techniques when studying mortgage default rates afflicted by compositional biases. And we believe that there are many ways to extend the simple hedonic models outlined in this paper to further enhance the accuracy of the insights.

1. Data

While S&P does not publish the underlying source data used for its SPIN Indices, these data pertain to all rated RMBS transactions issued in Australian dollars, which are covered by other data providers, including Bloomberg and Perpetual Trustees.

In this study we have relied on all Australian dollar RMBS deals available via Bloomberg, which are classified into "prime" and "non-conforming" in accordance with S&P's definitions. This data-set is rich in cross-section and on a monthly time-series basis, including detailed RMBS characteristic-level information, such as the monthly 90 day plus default rate, the number and value of the loans included in the pool, the weighted-average age of the loans in the pool, the weighed-average LVR at the time of issue, the date of issue, geographic exposures, types of borrowers, types of loans, and other data on the composition of the relevant portfolio.

[3] The Reserve Bank of Australia's data covers all home loans and is a much more stable sample than the smaller pool of loans included in the SPIN Index, the latter of which are more susceptible to material compositional biases as a function of large changes in issuance patterns. See the RBA publication "Household Indebtedness and Mortgage Stress" - Graph 7.

Although S&P and Bloomberg broadly cover the same transactions, there are some relatively trivial differences (e.g. deals that are not rated by S&P). As at January 2018 the total value of the unfiltered RMBS deals covered in the Bloomberg sample for the purposes of developing the hedonic indices was 102% of the value of the deals included in the relevant SPIN Index. This data-set was subsequently filtered to remove non-prime transactions and any pools with data quality concerns.

2. Method

Traditionally, hedonic methods have been used to monitor changes in price through regressions that adjust for compositional biases in observed samples of sale prices (e.g., monthly home sales or consumer price sales).[4] We believe that a hedonic regression method can be similarly employed to construct a mortgages delinquency risk index that adjusts for the abovementioned compositional biases.

The hedonic regression method assumes that heterogeneous goods can be described by their attributes. In a traditional house price index, the goods are houses; the statistic of interest is their intrinsic value; and observations are house sales. The asset attributes include the land area, the number of bedrooms and bathrooms, location, and other amenities, the distribution of which in the overall housing population is assumed to be constant across time periods. This distribution does, however, fluctuate greatly over time in the observed samples of sales. The hedonic regression method controls for these fluctuations in attribute distributions in the observed samples of sales to produce an index tracking the change in the overall housing market's intrinsic value in the absence of these biases.

In the context of mortgage default rates, the "goods" are mortgages, the statistic of interest is their delinquency rates, and the observations are monthly default rates of RMBS deals that are comprised of multiple mortgages. One core attribute of interest is months since the RMBS issue given the convention to structure RMBS transactions that are initially default free. While the distribution of this variable is irrelevant to the total mortgage population, it fluctuates materially across time in the sample of RMBS deals. Our hedonic index therefore aims to control for such fluctuations in the months since RMBS issue (amongst other variables) to produce a compositionally-adjusted index measuring the true change in mortgage delinquency rates.

There are two main approaches to calculating a hedonic index: the "imputation" and "time dummy" methods (see, for example, Diewert, Heravi, and Silver (2008) and CoreLogic's Hedonic Imputation Method (2018)). In the imputation approach, a regression is run for each time period and the index is constructed from the predicted prices based on the regression coefficients. An imputation method requires a database of the entire mortgage

[4] See, for example, CoreLogic's Hedonic Imputation Method for residential property prices.

population and all hedonic variables, which is not publicly available in Australia or for that matter any other developed world country. Accordingly, in this study we focus on the "time dummy variable" method.

To the best of our knowledge, hedonic regression methods have never before been applied to the problem of measuring mortgage default risk over time.

2.1 Time Dummy Variable Hedonic Model

The starting point[5] is the assumption that the delinquency rate d_n^t of RMBS deal n in period t is an additive function of a baseline rate δ^t at time t, plus a fixed number, say K, of characteristics of the deal as measured by z_{nk}^t. That is, we have the following indicative additive hedonic specification:

$$d_n^t = \delta^t + \sum_{k=1}^{K} \beta_k z_{nk}^t + \epsilon_n^t \quad (1)$$

where ϵ_n^t is a random error term. This is an indicative hedonic specification as in a later section, we describe the incorporation of non-linear hedonic variables.

We will conduct a pooled regression over a time window (nine years), such that the hedonic coefficients β_k are constant over this time window. In addition, as we are interested in underlying mortgages, but have only deal-level observations, we weight the regression by the dollar value of the loan pool in each deal w_n^t , denoting the total RMBS pool value at time t by $W^t = \sum_{n=1}^{N_t} w_n^t$ where N_t is the number of deals at time t.

While the final hedonic index is fitted via Restricted Maximum Likelihood methods (described in a later section), we will present the theory on a simplified case using ordinary least squares (OLS) regression, from which we can demonstrate valuable intuitions about the hedonic index. The OLS error function is as follows:

$$E = \sum_{t=1}^{T} \sum_{n=1}^{N_t} w_n^t \{d_n^t - (\sum_{t=1}^{T} \delta^t D_n^t + \sum_{k=1}^{K} \beta_k z_{nk}^t)\}^2 \quad (2)$$

where the time dummy variable D_n^t has the value 1 if the observation comes from period t and 0 otherwise. The baseline rates δ^t and hedonic coefficients β_k can be obtained via any standard statistic package. In order to gain intuition on this hedonic index, we will assume

[5] See also the Eurostat Handbook on Residential Property Prices (RPPIS) (2013).

the hedonic coefficients β_k are known and we mathematically solve for the time effects δ^t via differentiation by setting $\frac{\partial E}{\partial \delta^t} = 0$:

$$\begin{aligned}\delta^t &= \sum_{n=1}^{N_t} \frac{w_n^t}{W^t} d_n^t - \sum_{k=1}^{K} \sum_{n=1}^{N_t} \beta_k \frac{w_n^t}{W^t} z_{nk}^t \\ &= \overline{d^t} - \sum_{k=1}^{K} \beta_k \Big(\sum_{n=1}^{N_t} \frac{w_n^t}{W^t} z_{nk}^t \Big) \\ &= \overline{d^t} - \sum_{k=1}^{K} \beta_k \overline{z_k^t}\end{aligned} \tag{3}$$

Hence we arrive at the following intuition of the hedonic index: the change in the hedonic index $\delta^{t+1} - \delta^t$ is the change in deal-value-weighted average RMBS delinquency rate, plus hedonic adjustments:

$$\delta^{t+1} - \delta^t = (\overline{d^{t+1}} - \overline{d^t}) - \sum_{k=1}^{K} \beta_k (\overline{z_k^{t+1}} - \overline{z_k^t}) \tag{4}$$

The hedonic adjustment is the sum of each hedonic coefficient times the change in weighted-average hedonic characteristic. For example, let's say the hedonic coefficient β_m for the *months since RMBS issue* is positive, i.e. higher *months since RMBS issue* is correlated with a higher delinquency rate. If time $t+1$ has lots of new RMBS issues (i.e. many deals with low values of *months since RMBS issue*), the weighted average *months since RMBS issue* would be smaller, i.e. $\overline{z_m^{t+1}} - \overline{z_m^t} < 0$. This would lead to a positive hedonic adjustment, increasing the delinquency index δ^{t+1}.

An adjacent version of the hedonic index was also developed. We use a shorter pooling window size of 6 years, and slide this window forward one month at a time. For each new adjacent regression, we fix the time coefficients by taking the values from the previous adjacent regression (dropping the oldest one that fell off the sliding window). We therefore only estimate one new time coefficient for the most recent month. The other hedonic coefficients are allowed to vary across adjacent regressions to account for changing market conditions.

While the indicative hedonic specification was given in Equation 1, and the OLS regression solution given in Equation 3, we hereby describe additional modifications to these indicative methods. The hedonic specification in Equation 1 is purely linearly additive. However, we can incorporate non-linear additive terms in the hedonic specification by using spline decompositions, for example B-splines or P-splines which are smooth curves constructed by concatenating piecewise polynomial functions [6]. Instead of OLS, these linear and non-linear coefficients are estimated via the Restricted Maximum Likelihood Method.[7] The initial hedonic regression variables in our applied method are as follows:

- Non-linear variable for *months since issue* in view of the fact that all new RMBS transactions are initially calibrated to be default-free. Given extensive evidence of non-linear relationships between default probabilities and time,[8] we allowed this variable to be non-linear. Being a P-spline, it is penalised against overfitting;
- Weighted average LVR at deal origination as higher (lower) LVR loans have been empirically correlated with higher (lower) default rates[9];
- Weighted average loan age at deal origination as another control for the empirical relationship between time and default probabilities (noting that the weighted-average age of the loans and the age of the overall deal will have different influences on the probability of default).

Other variables could be included in the models in future research, such as those that control for the types of borrowers, the geographic composition of the pool, and the types of loans in the pool, all of which can exert influence on default risks.

It is worth highlighting that one key assumption in the usage of these hedonic variables is that their individual distributions in the total mortgage population are relatively constant over time. For example, we are assuming that the weighted average LVR in the total mortgage population is relatively constant over time. Extra care must be taken for variables for which this may not hold true.

2.2 Seasonal Adjustment

Inspecting the data we find that empirical default rates display clear seasonality: delinquencies typically increase after December (perhaps borrowers overspent at Christmas), and peak in June. To remove the seasonal component in our hedonic indices we use X-13ARIMA-SEATS (the latest seasonal adjustment software developed by the United States Census Bureau).

The SEATS approach models the input series with an ARIMA model, and then decomposes the obtained ARIMA model in the frequency domain into trend, seasonal and irregular components. (In the frequency domain the trend components will correspond to large

[6] See, for example: Eilers, P.H. and Marx, B.D., 1996. Flexible smoothing with B-splines and penalties. Statistical science, pp.89-102.
[7] Wood, S. and Wood, M.S., 2015. Package 'mgcv'. *R package version*, *1*, p.29.
[8] S&P's analysis implies that default rates peak approximately three years after the origination of a loan. See an Overview of Australia's Housing Market and Residential Mortgage-Backed Securities Market (2015).
[9] See, for example, Residential Mortgage Default Risk and the Loan-to-Value Ratio (2004)

values near zero frequency, the seasonal component will correspond to large values at regular frequency intervals and the irregular components correspond to a flat spectrum.) The seasonal components are adjusted in the frequency domain before retrieving the components in the time domain.

3. Results

Since we are developing hedonic indices, the absolute values of the index do not translate directly into a default rate—rather they measure compositionally-adjusted changes in the default rate over time. To facilitate comparison to simple weighted-average default indices, like the S&P SPIN Index, we rebase our hedonic indices to the SPIN Index level at the start of the relevant sample period.

Figure 1 shows the new hedonic index for RMBS defaults, which we denote the Coolabah Hedonic Index, rebased against the S&P SPIN Index for prime, public RMBS deals in Australia in both non-seasonally-adjusted and seasonally-adjusted terms. Whereas the SPIN and Coolabah Hedonic Indices correlate closely between 2010 and 2014, there is a notable disconnect between the two benchmarks after this time. Specifically, the seasonally adjusted Coolabah Hedonic Index documents an increase in the 90 plus day arrears rate from 0.55 per cent to over 0.65 per cent while the SPIN Index suggests the opposite. Significantly, this time period also saw a steady increase in the RBA reported arrears rate (Figure 1), which directionally reconciles with the Coolabah Hedonic Index changes even though the absolute levels are different because the RBA sample contains all bank balance sheet loans (and not just RMBS transactions). This time period also coincided with a large increase in RMBS issuance, which may have artificially depressed the SPIN Index even though the underlying or true rate of default was climbing. In Figure 1b one can also observe the impact of seasonal-adjustment, which removes the clear spikes in defaults that commence at the end of each year.

Fig. 1a

Delinquency Indices: SPIN, RBA vs Coolabah Hedonic Index

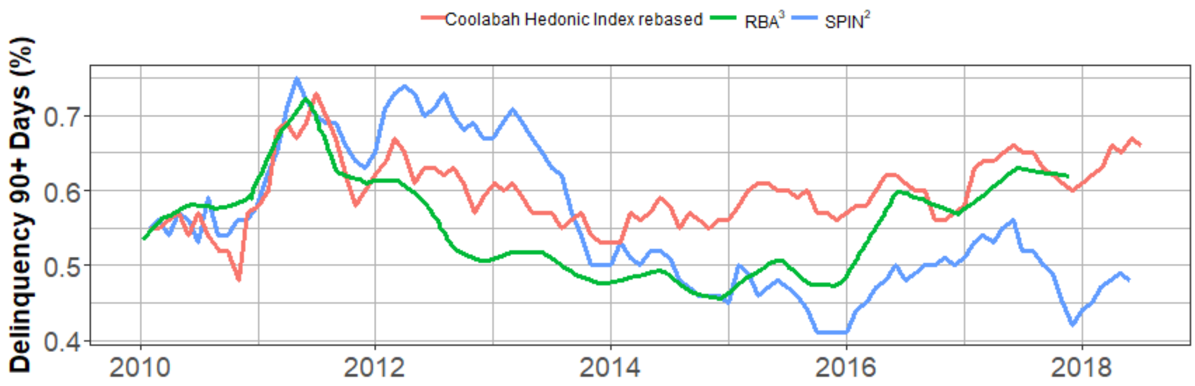
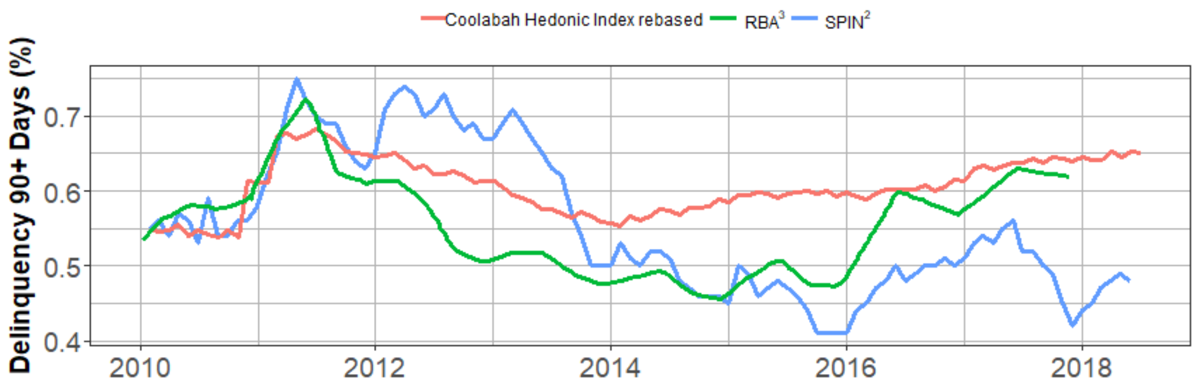


Fig. 1b

Delinquency Indices: SPIN vs Seasonally Adjusted Coolabah Hedonic Index



In a previous section, we established our intuition that changes in the hedonic index are driven by changes in both the weighted-average RMBS delinquency rates and the hedonic attribute variables. In Figure 2 one can see the changes in three characteristic variables: the weighted-average months since issue; the weighted-average age of the loans; and the weighted-average LVRs. In 2015, 2017 and 2018 there were noticeable declines in the months since issue variable (Figure 2a).

Yet perhaps the biggest attribute change was the decline in the average loan age over time, as highlighted by Figure 2b.

Fig. 2a
Pool Statistic: Weighted Average Months Since Issue

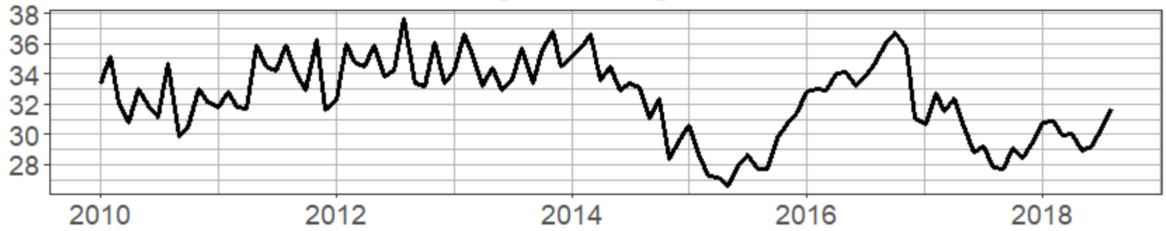


Fig. 2b
Pool Statistic: Weighted Average Collateral Age at Issue

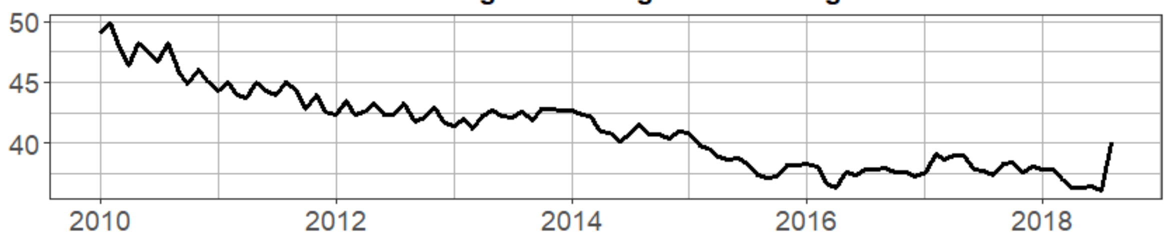
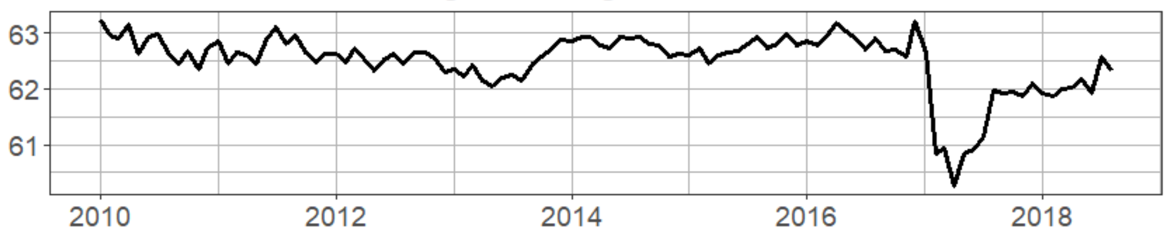


Fig. 2c
Pool Statistic: Weighted Average Loan-to-Value Ratio at Issue



To quantify these effects, we decompose the Coolabah Hedonic Index into the contributions from the modelled hedonic variables to visualise the impact of each hedonic adjustment. Figure 3 shows the weighted-average of our hedonic variables over time. As established previously (Equation 4), by multiplying these averaged values by negative one times their respective hedonic coefficients we can show the contribution from each variable over time (see Figure 3a). Summing up the contributions, we can add the summed adjustments to the raw RMBS delinquencies (with an arbitrary vertical shift) to reproduce the Hedonic Index as in Figure 3b. Note that there are minor differences due to the non-linearity we introduced in the *months since issue* hedonic variable, as well using the Restricted Maximum Likelihood parameter estimation method instead of OLS.

Fig. 3a
Hedonic Adjustment Components

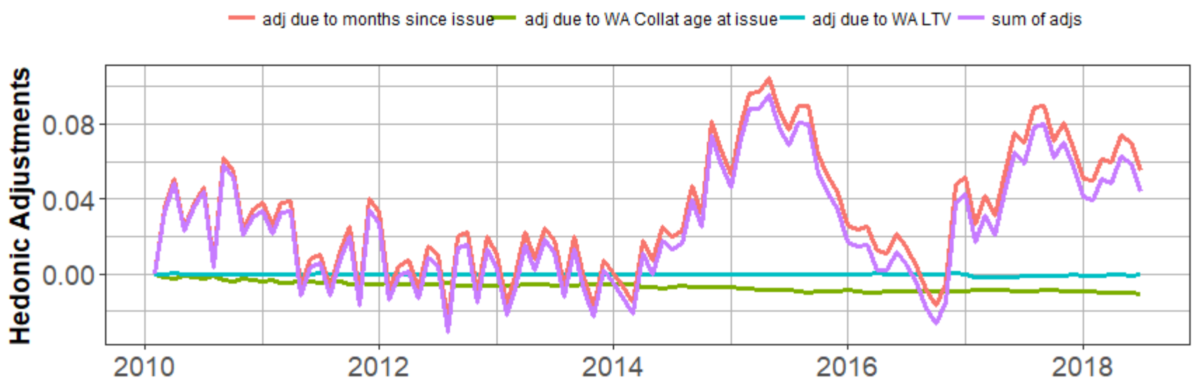
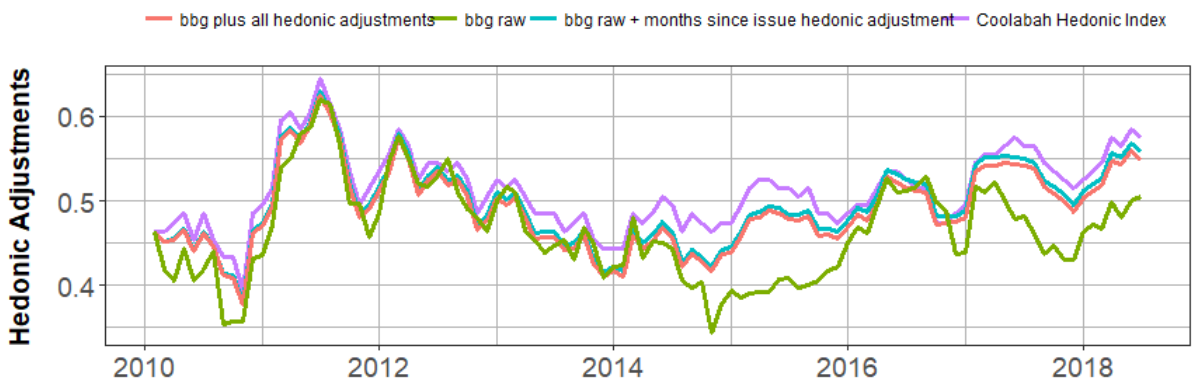


Fig. 3b
Effect of Hedonic Adjustment



4. Conclusion

In this paper we highlighted deficiencies with current market measures of mortgage arrears rates, specifically the compositional bias associated with RMBS delinquency rates exemplified by the S&P SPIN index.

To address the compositional bias problem, we developed the first known hedonic regression-based indices of default risk that explicitly control for compositional biases through the models' characteristic-based independent variables, which can include the amount of time since RMBS issuance, the weighted-average age of the loans, and the weighted-average LVR, amongst other factors.

Whereas the SPIN Index suggests that default rates have declined in recent years across Australian RMBS, our hedonic mortgage default index implies exactly the opposite: that is, compositionally-adjusted default rates on Australian RMBS deals have, in fact, been increasing sharply in recent times. Interestingly, our hedonic index reconciles with the arrears rates reported by the RBA, which are based on the performance of all bank balance-sheet loans (i.e. the population of all loans, which is not impacted by the aforementioned bias associated with changes in the volume and frequency of new RMBS issues).

Our findings have major implications for the RMBS market: participants' forward loss projections would have been overly optimistic if they are based on compositionally-biased indices, which they have been, instead of our compositionally-adjusted hedonic methods.

Going forward, we would encourage researchers to develop and use hedonic techniques when studying mortgage default rates afflicted by compositional bias. And we believe that there are many ways to extend the simple hedonic models outlined in this paper to further enhance the accuracy of the insights.