Applied Economics Letters

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/rael20

Measuring subjective housing affordability using a data-driven discrete information approach: A case study of Selangor, Malaysia

Jason Wei Jian Ng, Tomáš Želinský, Catherine S. Forbes & Cash Hao Looi

To cite this article: Jason Wei Jian Ng, Tomáš Želinský, Catherine S. Forbes & Cash Hao Looi (07 May 2023): Measuring subjective housing affordability using a data-driven discrete information approach: A case study of Selangor, Malaysia, Applied Economics Letters, DOI: 10.1080/13504851.2023.2208833

To link to this article: https://doi.org/10.1080/13504851.2023.2208833

APPLIED ECONOMICS LETTERS
https://doi.org/10.1080/13504851.2023.2208833

Routledge
Taylor & Francis Group

Measuring subjective housing affordability using a data-driven discrete information approach: A case study of Selangor, Malaysia

Jason Wei Jian Ng [a], Tomáš Želinský [b,c], Catherine S. Forbes [d] and Cash Hao Looi [d]

[a]Department of Applied Statistics, School of Mathematical Sciences, Sunway University, Bandar Sunway, Malaysia; [b]School of Education, Durham University, Durham, UK; [c]Faculty of Economics, Technical University Košice, Košice, Slovakia; [d]Department of Econometrics and Business Statistics, Monash Business School, Monash University, Melbourne, VIC, Australia

ABSTRACT

A widely adopted measure of housing affordability is that households should spend no more than 30% of their household income on housing. However, this normative threshold is an arbitrary Great Depression-era guideline and may not be relevant today. This paper proposes a subjective indicator of housing affordability by introducing a method commonly used in the medical sciences. It utilizes discrete information to estimate a subjective affordability ratio that discriminates between subjective house-poor and non-house-poor households. We apply the proposed method to household-level data collected in Selangor, Malaysia, and show that the optimal cut-off point is 23.5%. This estimated value suggests a higher prevalence of house-poor households than is implied by the regularly assumed 30% threshold. In addition, we perform a sensitivity analysis and find the bias in the estimated cut-off point is close to zero.

KEYWORDS

Subjective housing affordability; Youden index; Optimal cutpoint; Malaysia

JEL CLASSIFICATION

C43; G51; I31

I. Introduction

The 30% of household income standard as a gauge of housing affordability is still a widely-used Great Depression-era guideline that arose from the adage '*a week's wages to a month's rent*' (Pelletiere 2008). Nevertheless, this normative threshold has been widely criticized for being arbitrary (Baker, Mason, and Bentley 2015) despite there being no consensus on the appropriate value (Bramley 2012), with percentages used in the literature varying between 25% and 50% (Chen, Hao, and Stephens 2010).

The increased use of subjective indicators in social sciences has prompted work measuring perceived affordability (Heylen 2021), also known as subjective housing affordability. Growing evidence shows that normative affordability thresholds may not be congruent with people's perceptions of housing affordability. Significant disparities between objective and subjective indicators of socioeconomic conditions can lead to inefficient allocation of public resources and weaken trust in official statistics (Stiglitz, Sen, and Fitoussi 2009). To address this issue, Heylen (2021) proposes a variable threshold that better reflects perceived housing cost burdens. Sunega and Lux (2016) showed that alternative threshold definitions may decrease the significant discrepancy between subjective perceptions and objective indicators of housing affordability. Despite efforts to redefine the threshold, most methods still seem arbitrary. Therefore, we propose a data-driven approach to determine the optimal threshold that accounts for both objective and subjective data.

Our main objective is to introduce a method commonly used in the medical sciences utilizing discrete information to estimate this subjective affordability ratio (SAR) to discriminate between subjectively reported house-poor and non-house-poor households. This method has also been recently applied in the material deprivation literature (Želinský, Mysíková, and Ng 2021). We illustrate the proposed method using household-level data collected in Selangor, one of Malaysia's key economic states.[1]

CONTACT Tomáš Želinský tomas.zelinsky@durham.ac.uk School of Education, Durham University, Stockton Road, Durham DH1 3LE, UK

[1]Selangor is one of Malaysia's highly developed states with the largest urban population of 6.7 million. This translates into an urbanization rate of 95.8%. Selangor's median house price as of Q2 2022 is RM430,000, the third most expensive state in the country (NAPIC 2022).

Supplemental data for this article can be accessed online at https://doi.org/10.1080/13504851.2023.2208833.

II. Methodology and data

Measuring SAR

A common approach to measuring subjective housing affordability involves households' *self-reported* financial burden levels – discrete information based on perceptions of their housing costs (Özdemir Sarı and Aksoy Khurami 2018). We augment this information by incorporating data on households' *actual* housing expenses as a percentage of household income. In determining their reported degree of burden, individual households are assumed to compare their actual housing expenses to a latent housing affordability threshold z. Households will self-assess themselves as house-poor if their housing expense exceeds their z value.[2] However, as households will have varying housing expenditure ratios and values of z, our objective is to find the optimal threshold that minimizes the classification discrepancies between actual expenditure ratios and the reported perceived financial burden groups. Specifically, we aim to minimize the number of households that report feeling house-poor despite having actual expenditure ratios below the estimated threshold (false house-poor), and the number of households that do not report feeling house-poor despite having ratios above this threshold (false non-house-poor). Figure 1 illustrates this approach. We estimate the value of z^* by maximizing the Youden index.

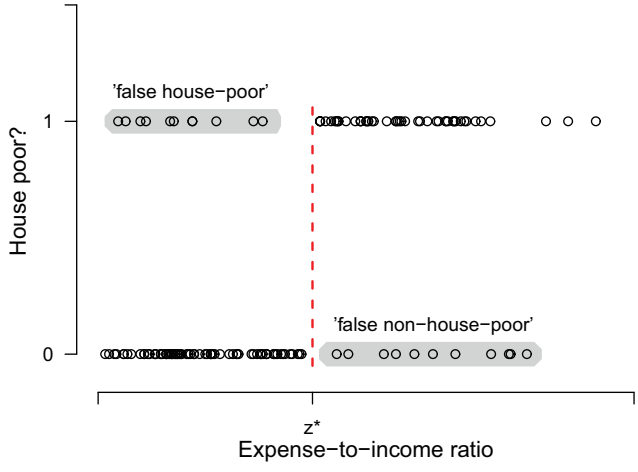

Figure 1. Estimating a subjective housing affordability ratio with discrete information.

Youden index

The Youden (1950) index was originally developed in medical research to determine how accurately a diagnostic test can distinguish between patients with a disease and those without. The test relies on a cut-off point c, which groups patients based on their biomarker values. In this study, we use the Youden index to identify the cut-off point z, differentiating between households that feel house-poor and those that do not. The index, denoted by J, is calculated using sensitivity and specificity measures for all possible values of z, such that

$$J(z) = \{Se(z) + Sp(z) - 1\}.$$

At a given z, (Se) and (Sp) represent the probabilities of correctly classifying subjective house-poor and non-house-poor households, respectively, with the value of $J = 1$ indicating perfect classification and $J < 1$ otherwise (Shan and Rapallo 2015). The *optimal* cutpoint z^* is the one that maximizes J, ensuring that the greatest number of households are correctly classified.

Data

In March 2022, a survey was conducted among 1,211 Selangor households using a proportional sampling scheme that included 70% homeowners and 30% renters. To ensure the representativeness of the data, a marketing consulting firm in Malaysia was engaged to collect the data via self-administered questionnaires distributed through their online panel. The self-developed questionnaire included questions on specific housing expenses, such as the monthly amount spent on mortgage or rent, utilities, maintenance, house insurance, and net household income and demographic characteristics. The actual housing expense-to-income ratio was immediately calculated from the responses to these questions and displayed to the respondents who were then asked to indicate the level of financial burden they perceived from their housing expenses (heavy, slight or none). Further details of the survey can be found in the supplementary material.

[2]'House-poor' households refer to households that *perceive* their housing costs to be a financial burden and hence, unaffordable.

As the Youden index is a binary classifier, we dichotomize the subjective indicator by only referring households as subjectively house-poor if they reported their housing expenses as *a heavy burden*.

III. Results

Estimated SAR

The results indicate the optimal cutpoint z^* discriminating between subjectively house-poor and non-house-poor households is 23.5%, implying that households spending more than 23.5% of their household income on housing expenses consider themselves as house-poor. This threshold, based on respondents' perceptions of housing burden, is lower than the commonly used value of 30%. However, the 95% confidence interval for z^* (22.9, 36.0) produced using a bootstrap procedure is wide and includes the standard 30% value, as presented in column (2) of Table 1.

The distribution of housing-expenses for respondents who reported '*a slight burden*' (middle category) has considerable overlap with the other categories ('*a heavy burden*', '*no burden at all*'), as seen in Figure 2. This overlap adds uncertainty to the estimation procedure, a fact further illustrated in the left panel of Figure 3 where the bootstrap distribution of optimal cutpoints exhibits a second large peak around 27%.

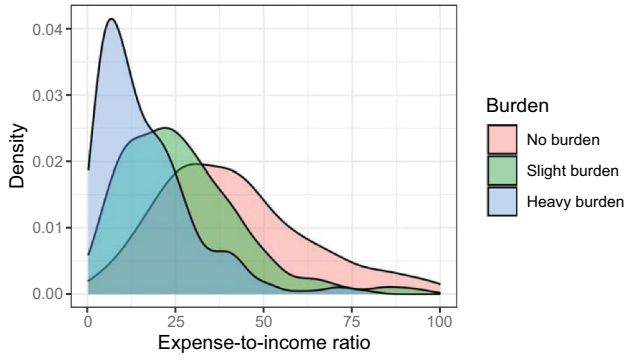

Figure 2. Distribution of housing related expenditure, by subjective financial burden indicator.

The middle category corresponds to respondents who perceive their housing-related expenditures to be '*a slight burden*'. However, the middle category may represent a neutral opinion, where the respondent either has no strong opinion, or does not wish to reveal a strong opinion, in relation to the two more extreme choices (Kulas and Stachowski 2009). To evaluate the robustness of our primary finding to the middle category, we also examined the results by excluding cases where respondents selected this category. As shown in column (3) of Table 1, the optimal cutpoint $z^* = 23.45\%$ was estimated when excluding cases where the middle category was indicated, a value very close to and not significantly different from our original estimate of 23.46% ($p = 0.997$). Furthermore, the 95% confidence interval associated with the estimate that excludes the middle category, as shown in column (4) of Table 1, is narrower than the original interval. This result is supported by the right panel of Figure 3, which displays a distinct global mode at approximately 23.5%.

Our estimated optimal cutpoint based on maximizing the Youden J index corresponds to the value $J = 0.39$ when using all responses and $J = 0.59$ when excluding the middle self-assessed financial burden category. As an index of $J = 1$ indicates perfect classification, our obtained values may raise concerns regarding potential bias associated with the estimated optimal cutpoint. Therefore, we conduct a sensitivity analysis to assess the degree of bias in the maximized Youden J index.

Table 1. Estimated cutpoints and associated characteristics.

	Heavy vs Slight/No financial burden		Heavy vs No financial burden (Slight burden excluded)		Diff (3) vs (1)
Metric	Point est. (1)	Bootstrap CI (2)	Point est. (3)	Bootstrap CI (4)	p-value (5)
Cutpoint z* (SAR)	23.46	(22.93, 35.96)	23.45	(20.60, 28.65)	0.997
Youden $J(z^*)$	0.387	(0.345, 0.454)	0.590	(0.531, 0.658)	<0.001
Sensitivity	0.814	(0.582, 0.850)	0.814	(0.707, 0.866)	>0.999
Specificity	0.573	(0.552, 0.806)	0.776	(0.719, 0.873)	0.008

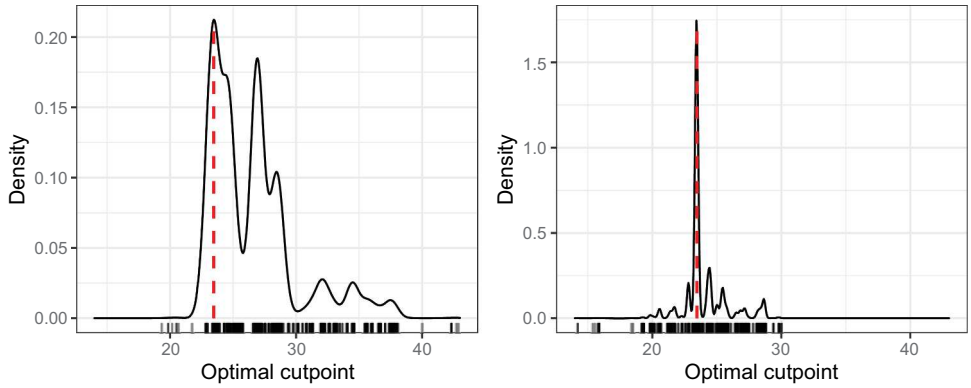

Figure 3. Bootstrap distribution of optimal SAR for 2 categories; Left panel: Heavy vs slight or no burden, right panel: heavy vs no burden (middle category removed).

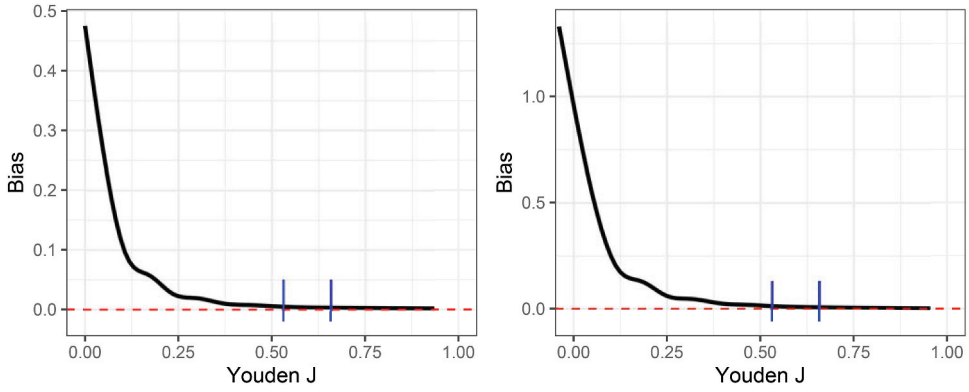

Figure 4. Sensitivity analysis: Average relative bias against Youden J; Left panel: Heavy vs Slight or No burden, Right panel: Heavy vs No burden (middle category removed).

Sensitivity analysis

We vary the true SAR starting with the lower bound of the bootstrap interval estimate given in columns (2) and (4) of Table 1, incrementing by 0.05 until the upper bound is reached. In addition, we randomly vary the proportion of subjectively house-poor respondents according to a binomial probability ranging from 0.05 to 0.95 in increments of 0.05 to introduce noise in the data. We use 100 replications for each scenario and compute the (average relative) bias as the difference between the estimated and true SAR, divided by the true SAR, with zero indicating no bias. The results, shown in Figure 4, indicate that the bias drops considerably to below 0.025 (0 = no bias) when $J > 0.25$. The blue vertical lines in the left and right panels of Figure 4 indicate interval estimates of the Youden values (i.e. $J(z^*)$) reported in columns (2) and (4) of Table 1, respectively. In both cases, it can be seen that the bias is close to zero.

IV. Conclusion

Our study offers a new method of estimating subjective housing affordability based on the Youden J index and a subjective financial burden classification. Applying the method to survey data from Selangor, Malaysia, we found evidence that the cutpoint distinguishing the house-poor from the non-house-poor is 23.5%, significantly lower than the normative 30% threshold usually adopted in Malaysia. Moreover, the results of our sensitivity analysis support the

claim that the estimated optimal cutpoint is unbiased. The implication of our findings is that more households are struggling with housing affordability than the official threshold of 30% would suggest.

Our approach can serve as a useful reference for researchers or policymakers seeking to apply similar methods in their specific context, while also highlighting the importance of performing sensitivity analyses to avoid the adoption of biased estimates.

We have demonstrated the robustness of our results by excluding the middle category and obtaining a more precise cutpoint estimate. Future research will explore the use of multiple optimal cutpoints estimated from multi-categorical discrete information.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The work was supported by the Monash University Faculty of Business and Economics Australia-Malaysia 2021-2022 Research Collaboration Development Scheme, Grant code STG-000088. Forbes also acknowledges financial support by the National Science Foundation (NSF) Grant SES-1921523. Želinský also acknowledges financial support by the Slovak Scientific Grant Agency (VEGA) Grant 1/0034/23. The responsibility for all conclusions drawn from the data lies entirely with the authors.

ORCID

Jason Wei Jian Ng http://orcid.org/0000-0001-6832-6267
Tomáš Želinský http://orcid.org/0000-0001-7198-0278
Catherine S. Forbes http://orcid.org/0000-0003-3830-5865
Cash Hao Looi http://orcid.org/0009-0002-7219-2753

Ethical approval

Ethics approval for this study was obtained from the Monash University Human Ethics Committee (Project ID: 31276).

References

Baker, E., K. Mason, and R. Bentley. 2015. "Measuring Housing Affordability: A Longitudinal Approach." *Urban Policy and Research* 33 (3): 275–290. doi:10.1080/08111146.2015.1034853.

Bramley, G. 2012. "Affordability, Poverty and Housing Need: Triangulating Measures and Standards." *Journal of Housing and the Built Environment* 27 (2): 133–151. doi:10.1007/s10901-011-9255-4.

Chen, J., Q. Q. Hao, and M. Stephens. 2010. "Assessing Housing Affordability in Post-Reform China: A Case Study of Shanghai." *Housing Studies* 25 (6): 877–901. doi:10.1080/02673037.2010.511153.

Heylen, K. 2021. "Measuring Housing Affordability. A Case Study of Flanders on the Link Between Objective and Subjective Indicators." *Housing Studies* 38 (4): 552–568. Advance online publication. doi:10.1080/02673037.2021.1893280.

Kulas, J. T., and A. A. Stachowski. 2009. "Middle Category Endorsement in Odd-Numbered Likert Response Scales: Associated Item Characteristics, Cognitive Demands, and Preferred Meanings." *Journal of Research in Personality* 43 (3): 489–493. doi:10.1016/j.jrp.2008.12.005.

National Property Information Centre (NAPIC). 2022. "Property Market Price & Rental Indices Q2 2022." *Valuation and Property Services Department*. Accessed 23 February 2023. https://napic.jpph.gov.my/portal/key-statistics

Özdemir Sarı, Ö. B., and E. Aksoy Khurami. 2018. "Housing Affordability Trends and Challenges in the Turkish Case." *Journal of Housing and the Built Environment* 38 (1): 305–324. doi:10.1007/s10901-018-9617-2.

Pelletiere, D. 2008. *Getting to the Heart of Housing's Fundamental Question: How Much Can a Family Afford? A Primer on Housing Affordability Standards in US Housing Policy*. Washington, DC: National Low Income Housing Coalition.

Shan, G., and F. Rapallo. 2015. "Improved Confidence Intervals for the Youden Index." *PloS One* 10 (7): e0127272. doi:10.1371/journal.pone.0127272.

Stiglitz, J. E., A. K. Sen, and J. -P. Fitoussi. 2009. *Report by the Commission on the Measurement of Economic Performance and Social Progress*. http://www.stiglitz-sen-fitoussi.fr/documents/rapport_anglais.pdf

Sunega, P., and M. Lux. 2016. "Subjective Perception versus Objective Indicators of Overcrowding and Housing Affordability." *Journal of Housing and the Built Environment* 31 (4): 695–717. doi:10.1007/s10901-016-9496-3.

Youden, W. J. 1950. "Index for Rating Diagnostic Tests." *Cancer* 3 (1): 32–35. doi:10.1002/1097-0142(1950)3:1<32:AID-CNCR2820030106>3.0.CO;2-3.

Želinský, T., M. Mysíková, and J. W. J. Ng. 2021. "Choosing an Optimal Material Deprivation Indicator Threshold." *Applied Economics Letters* 28 (2): 100–104. doi:10.1080/13504851.2020.1734181.