University Library



A gateway to Melbourne's research publications

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
O Flaherty, B;Scutella, R;Tseng, YP

Title:
Using Private Information to Predict Homelessness Entries: Evidence and Prospects

Date:
2018

Citation:
O Flaherty, B., Scutella, R. & Tseng, Y. P. (2018). Using Private Information to Predict Homelessness Entries: Evidence and Prospects. Housing Policy Debate, 28 (3), pp.368-392. https://doi.org/10.1080/10511482.2017.1367318.

Persistent Link:
https://hdl.handle.net/11343/197669

Using Private Information to Predict Homeless Entries: Evidence and Prospects

Abstract

Do people at risk of homelessness have private information—information that social service agencies cannot credibly obtain—that helps predict whether they will become homeless? Our answer is yes: homeless people and people at risk of homelessness know important things about their future. We use data from Journeys Home, a path-breaking longitudinal study of people experiencing homelessness and people at risk of homelessness in Australia. In many cases, the private information that participants have predicts entries better than the public information that agencies can obtain. We suggest ways that private information can be used to improve service delivery.

* This paper uses data from the Journeys Home study, which was initiated and is funded by the Australian Government Department of Social Services (DSS), and is managed by the Melbourne Institute of Applied Economic and Social Research (Melbourne Institute). The findings and views in this paper should not however be attributed to either the DSS or the Melbourne Institute. We are grateful to Martha Burt, Pierre-Andre Chiappori, Abraham Chigavazira, Dennis Culhane, Beth Shinn, participants at the 2016 Workshop on Homelessness and Housing Insecurity at the Melbourne Institute, and members of the National Alliance to End Homelessness (United States) research council for helpful comments and questions. We also thank two anonymous referees for major improvements in the manuscript.

Keywords: homelessness, homeless entries, homelessness prevention, private information, longitudinal studies

1 Introduction

Homelessness prevention programs try to target their services to households who would soon become homeless in the absence of intervention, but targeting requires information. This paper studies who has relevant information about impending entries into homelessness, and how program operators can harness that information for better targeting.

Some relevant information is either publicly available or easy for program operators to obtain from households. This is the sort of information that intake questionnaires usually collect, and it often has some predictive power about future events: very rich households, for instance, are not likely to become homeless very soon. We call this "public information."

In addition to public information, households generally have a lot of other information that may be relevant, but this information may be impossible to collect through a questionnaire. There are many different reasons why program operators cannot obtain such information from households through a questionnaire. The information may be unverifiable and some households may not answer a question truthfully if the answer mattered to how they would be treated ("how badly do you need money?"). The information may be idiosyncratic so that formulating the relevant questions is not feasible ("do you have a cousin in Puerto Rico who is planning to move to New Zealand and leave you a place to stay?"). The information may be subjective and the household may not have a good metric for expressing it ("how embarrassed would you be to enter a shelter?") or may be ashamed to answer the question ("did your father sexually abuse you when you were little?"). Or the program operator may be prohibited by law from using some information to decide on which services if any a household receives (in the United States, "what is your race?" cannot be used for these purposes). (More rigorous definitions of public and private information are provided in section 3.)

In this paper, we ask whether people at risk of homelessness have useful private information about whether they are likely to become homeless in the near future. Our answer is yes. We use data from Journeys Home (JH), a longitudinal study of people at risk of homelessness in Australia.

Thinking about private information also helps in understanding how to tackle many other program design problems in homeless services. For instance, a companion paper explores the question of whether homeless people have useful private information which would indicate when they will exit homelessness. This question helps in designing rehousing programs. Thus, the current paper has wider applications.

Why does a consideration of homeless entry matter? Programs that engage participants' private information about the future have rarely been deliberately used in the field of homelessness, but they are common elsewhere. Such programs generally generate better outcomes than those that restrict themselves to public information.[1] But designing programs to incorporate private information makes sense only if participants have such relevant information.

We can show that people at risk of homelessness have relevant private information because the Journeys Home longitudinal survey asks participants at regular intervals over time for subjective evaluations of their current housing situations and forecasts of changes in the near future. Their answers to questions in JH have no imminent personal consequences because JH is a survey and their privacy is protected. Holding a rich vector of public information constant, the survey responses significantly affect the probability of becoming homeless in the next period. The significance is both economic and statistical.

The private information that people at risk of homelessness have opens up many opportunities for understanding how existing programs work and how new programs can be designed. In this paper we concentrate on homelessness prevention.

The plan of the paper is as follows. The next section reviews the literature on homelessness prevention and explains how many of the practices that we see in normal life implicitly use private information; the approach that we use is not exotic (although some of the language may be new to

[1] A large literature in mechanism design gives many examples. Well-known examples range from optimal taxes (Mirrlees 1971) to assignment of residents to hospitals (Roth 1990) to award of electromagnetic spectrum licenses (Fox and Bajari 2013).

homeless services). From auctions to soup kitchens, when people want to find out other people's private information they usually use methods that are much more adept than questionnaires and intake forms. Section 3 describes the data and the methods we use. Here the goal is to find out how well the private information that people at risk of homelessness have predicts imminent homelessness. Section 4 presents the main results: private information improves the predictions possible with public information considerably. In section 5 we turn to applications and discussion, and simulate how various program design operations would affect who seeks to participate in prevention programs and how the resulting applicant pool would allow program operators to deliver services. We also examine the US Veterans Administration universal screener, a tool that elegantly retrieves private information and uses it well. If private information is indeed the "secret sauce" that makes prevention work, then we want to show some of the different ways it can be mixed.

2 Literature review and general background

2.1 Homelessness prevention

Recent evaluations of homelessness prevention programs in New York City (Rolston et al, 2013, Goodman et al., 2016) and Chicago (Evans et al., 2015) have shown favorable ratios of benefits to costs. These results were not completely anticipated: experts in the early 21st century generally felt that prevention was not a promising way to combat homelessness (Shinn et al, 2001, Burt et al, 2005). The reason for this is that targeting was considered too hard because many households looked too similar on observable characteristics to those who became homeless in any short period. Thus finding the right households to treat looked like an almost impossible task.

Why then did the programs succeed? One plausible story is that the right households "found" themselves, instead of waiting for program operators to find them. They acted on their private information indicating that they were likely to become homeless in order to show up and ask for prevention services.

Consider New York City. During six months around the time of the relevant evaluation study

about 2.7 percent of New York's poor households with children entered homeless shelters.[2] But in the randomized control trial conducted in 2010, 14.5 percent of the control group—families who sought prevention services but were randomly chosen not to receive them—entered shelters within six months. The families who sought prevention services were more than five times more likely, in the absence of services, to become homeless than the average poor family. If they had been average poor families, the program would not have been successful.

What made the families who sought services different from those who did not? One reasonable possible explanation is that they thought they were likely to become homeless; that is, they had valuable private information. By showing that people at risk of homelessness have valuable private information we support this explanation of why some homelessness prevention programs are successful, and facilitate an understanding of the challenges of designing successful prevention programs.

2.2 Private information and program design

Vickrey (1961) is the earliest study about designing mechanisms to elicit private information, and using it to improve allocation. Vickrey studied auctions.

Suppose a seller is trying to dispose of a valuable painting. There are many potential buyers and each potential buyer has a reservation price, the maximum amount she is willing to pay for the painting. If the seller knew the reservation prices, she could go to the buyer with the highest reservation price, make a take-it-or-leave-it offer at that price, and collect the highest reservation price. But reservation prices are the private information of potential buyers, and the seller does not know them.

One way for the seller to proceed is to stage an English auction. An English auction will induce

[2] In 2007-2011, about 225,000 families with children lived in poverty in the city (American Community Survey), and about 1000 families a month entered shelters (Goodman et al, 2016). So in six months about 6000 families would enter shelters. Since some families may enter twice, since there is turnover among poor families, and since some families not previously resident in New York may enter shelters, the estimate of 2.7 percent is probably high.

the second highest valuer among the potential buyers to reveal his reservation price, and the highest valuer to reveal who she is. That is because the second highest valuer will drop out of the auction when the price rises above his reservation price, and the highest valuer will remain and win the auction. So an auction is a way to get potential buyers to reveal their private information: the seller does not give questionnaires to the potential buyers and try to predict based on their childhood experiences which one will value the painting the most.

The seller, however, must bear a cost for this information revelation: she cannot collect the full value of the painting to the highest valuer, only the value to the second highest valuer.

Nichols et al. (1971) showed that being asked to wait can cause people to reveal private information about the value of their time. Nichols and Zeckhauser (1982) observed that providing in-kind services rather than cash was also a way of inducing households to reveal their information. This insight appears in an informal way in many practices that homeless service providers use.

Consider, for instance, the use of in-kind services rather than cash—soup kitchens and shelters, for instance, rather than payments to recipients that would allow them to buy these items themselves. This practice is particularly prevalent with charitable organizations that have no ability to perform means-testing. One reason for this is private information revelation: the charities do not know the resources of potential recipients, and by offering a meal or a cot they find out which potential recipients are so poor that they are willing to accept the meal or the cot. Money would not provide the same information because everyone would want the money. Notice that, once again, information revelation comes at a cost: the charities may provide meals or beds of lower quality than they might otherwise wish in order to keep themselves from being inundated by people less needy than those they are trying to reach.

In an empirical paper, Alatas et al. (2016) extend the insight about waiting time to inconvenience in general and provide a particularly instructive example of how small changes in program accessibility can induce people to reveal private information in a way that improves targeting. They studied a conditional cash transfer program in Indonesia, randomly assigning

different application procedures to different villages. The program reduced the poverty gap much more in villages where potential participants faced greater difficulties in enrolling. In those villages, richer households were discouraged from applying by the time involved andbecause they thought they were less likely to be accepted. In the villages where application was easier, rich households applied anyway, and, because the review process was imperfect, some of them were accepted.

Thus designing programs that use recipients' private information is not unusual and is even common in homeless services practice. The private information most commonly used, though, is contemporaneous—what resources are available right now. The question we ask is whether homeless people and people at risk of homelessness have relevant private information that predicts the near future, not just the present.

Our answer in section 4 is that they do. Then in section 5 we approach prevention programs in much the same spirit that Alatas et al (2016) approached conditional cash transfer programs in Indonesia and show, as they do, that the most effective programs are probably not the most convenient. But we use simulation rather than a randomized control trial.

3 Data and methods

3.1 Data

We use data from the Journeys Home (JH) survey, an interviewer-administered survey that followed a representative sample of Australian welfare recipients facing homelessness or housing insecurity between September 2011 and May 2014. Participants were assured that participation was entirely voluntary and that the information that they provided would be kept strictly confidential and for research purposes only.[3] For further details of the survey methodology and sampling strategy see Wooden et al (2012).

[3] For evidence of this see the information brochures all participants received at: http://melbourneinstitute.com/journeys_home/participate/publications.html

We use all six waves of the survey, with the Wave 1 interview conducted between September and November 2011 and the five additional follow-up interviews carried out at six-month intervals. As discussed in Wooden et al (2012) and Cobb-Clark et al (2016), initial response and re-interview rates compare favourably with other nationally representative longitudinal studies.

As the initial JH sample comprised people who were either homeless or vulnerable to homelessness at the time the sample was drawn, over the course of survey entries to homelessness were observed for many members of the sample, whatever their status at the outset. For many not initially homeless, homelessness occurred later on. Also for some originally homeless individuals, repeat episodes occurred later on. This, coupled with the non-consequential nature of survey responses, makes the JH data ideal for our analysis.

3.2 Methods

We want to know whether people at risk of homelessness have useful private information about whether they are likely to become homeless in the near future by examining entries into homelessness (the type of information that a homelessness prevention program would care about). Specifically, we fit probit equations of the form:

$$pr(h_{it} = 1) = \alpha U_{it} + \beta R_{it} + \varepsilon_{it}.$$

where $h_{it}$is an indicator variable equal to 1 if, and only if, household i becomes homeless in the three-month period following interview t conditional on not being homeless at t (which includes both those never homeless prior to t and those previously homeless who exited homelessness prior to t) ; U_{it} is a vector of public information about household i at interview t; R_{it} is a vector of private information about household i at interview t; and α, β are vectors of coefficients. The major question for this paper is whether we can reject the null hypothesis that the vector β is the zero vector (all its components are zero). The three-month time horizon is long enough for transitions to be common, but short enough that the information might be useful for a homelessness prevention program. Further detail on how homeless entries are defined and on sample selection is provided in Appendix A.

Essentially, we first find out how well all the public information together predicts homeless entries. Then we add private information to see if private information makes the predictions more accurate. If so, then we can say that private information would be useful to service providers if they could obtain it (which they cannot do by just asking questions). More practically, if people at risk of homelessness have useful private information, then service providers may learn by observing how these people act in certain circumstances (just as auctioneers learn something useful by observing who stops bidding).

3.3 Measures of public information

By "public information" we mean all information that a service provider can reliably gather from an applicant herself or from some trusted secondary source, and which the service provider can legally use in determining which services, if any, to offer the applicant. Hence what information is public varies by time, place and technology.

Operationally, we base our choice of variables on the Australian Specialist Homeless Services Initial Client Form (see Table 1 for a list of variables used and Table A1 in Appendix A for full variable definitions and summary statistics). Of course, the intake form does not exhaust the information that the Australian Specialist Homeless Services could have gathered—for instance, it could have asked for applicants' shoe sizes. But we assume the absence of such questions from the form indicates a belief that such information would not be useful in determining services to be provided.

On the other hand, it is also possible that applicants may answer some questions on the intake form wrongly in ways that cannot be verified. For instance, consider participants who tell JH that they have a diagnosed mental health condition but wrongly tell Australian Specialist Homeless Services that they don't. If Australian Specialist Homeless Services uses the JH public information equation to predict who among the people it sees will be homeless, then the Australian Specialist Homeless Services prediction will be less accurate than we measure it to be. (If Australian Specialist Homeless Services do not use the JH public information equation, their prediction will

also be less accurate.) If our estimate of the accuracy of the public information equation is an overestimate of what program operators experience for this reason, we will be biasing our results towards a finding of no relevant private information, given public information.

For both public and private information, we use a very large number of variables. We expect that many of them are collinear with other variables and that many will not have significant coefficients in the regressions we will run. This is not a concern, as we e are interested in finding out how well these large groups of variables jointly predict, not so much in how the individual variables predict. We want to find out what the "team" of private information variables does, not what the individual "players" do.

3.4 Measures of private information

By "private information" we mean information that an applicant has that is not public, as defined in section 3.3. So private information includes information that the applicant knows but might not be willing or able to express to the service provider if the answer were consequential, information that the service provider cannot legally use, and information so idiosyncratic that service providers do not ask about it.

Operationally we take private information to be answers to questions in JH that do not correspond to questions in the Australian Specialist Homeless Services Initial Client Form. Respondents' answers to these questions in JH have no personal consequences because JH is a survey and their privacy is protected. Clearly, these questions do not exhaust all the private information that applicants might have. JH lacks detailed information about job prospects, places to live or opportunities for family reconciliation, as well as information about preferences, character or talents. To the extent that this information is not reflected in JH measures we use, we bias our results against a finding that private information matters.

It is important to our project that the answers be non-consequential, not that they be true. By non-consequential we mean that no way of answering the question will make the applicant better off or worse off. Suppose JH included as a question, "What is your favorite color?" Suppose

also that people who answered “yellow” were much more likely to become homeless, holding all other variables constant, than people who answered other colors. The prediction by itself is sufficient to establish the importance of private information (although not necessarily private information about liking the color yellow) for two reasons.

The first reason is that this relationship does not imply that the question could be imported to a consequential intake questionnaire with similar results. If people who said “yellow” on an intake questionnaire were treated more favorably or less favorably as a result because they were believed to be more likely to become homeless, then an unknown fraction would change their answers and the original relationship between “yellow” and homelessness would dissolve. The answer to the favorite color question cannot be treated like an answer to a question about public information.

Second, the answer tells us about private information even if it is not true. Something makes people who answered “yellow” different from people who named other colors. That something is not reflected in public information, and that something predicts homelessness. Hence that something is private information. For instance, if people answer “yellow” because they like the sound of the word, whether they like the color or not, then people who like the sound of the word “yellow” are more likely to become homeless, and fondness for the sound of the word “yellow” is private information. So truth of answers in JH is not relevant to our purpose (though it would be, for instance, if we were trying to test a theory about causes of homelessness).

A full list of private information variables used is provided in Table 1.We use three sets of variables.

The first set is what we call direct private information variables. These variables are answers to two questions that directly capture people’s assessments of the security of their housing arrangements. Specifically, they are answers to the following questions:

a. ‘Can you sleep here for the next 3 months without being asked to leave?’ We call this the “continuation question.”

b ‘Which one of the following best describes how you see your housing situation at the

moment?'

- Secure in your accommodation;
- At risk of being homeless;
- Homeless; or
- Don't know/Refused.

We call this the "housing situation question."

Notice that people may use some idiosyncratic information in formulating their answers to the direct housing questions. For instance, if a person has a great relationship with his current landlord, that information may be reflected in his answer to the question about the next three months. To that extent, the omission from JH of a question like "How do you get along with your current landlord?" is not a problem and does not bias down the estimated impact of private information. But these questions are not broad enough to reflect private information like job prospects or family reconciliation that probably do affect homeless entries.

The second set of private information variables is what we call indirect housing information. These variables are answers to questions about housing, but not its security. These are questions about overcrowding; perceptions of safety, adequate sleeping space and general good condition; obstacles to finding another place; and general satisfaction with their current housing situation.

Indirect housing variables also include information about homeless history: for example, whether the respondent has a history of long-term homelessness or of sleeping rough. Because in most cases this information cannot be verified, we consider it private information. In some circumstances however—for instance, long shelter stays in New York City—this information can be verified, and so would be considered public rather than private. Our basic results change very little if we treat homeless history as public rather than private information.

The final set of private information variables are not related to housing, but may predict homelessness. We call these non-housing variables. These are answers to questions about incarceration or detention, physical or sexual violence in the last six months, debt, drug injection,

risky behavior, social support, and a series of questions about whether lack of money forced the respondent to pawn or sell something, go without food while hungry, ask a welfare agency for money or in-kind assistance, ask friends or family for help, miss going out with friends, or delay paying utility bills.

Notice that most of the indirect housing variables and indirect variables are not forward-looking; they ask about current or past conditions, not anticipations of future conditions. Knowing about their impact on future homelessness is important, however, because they might predict future homelessness and JH contains no direct prediction equation like "How likely is it that you will be homeless in three months?" If JH contained such a question, we would not use the indirect housing and indirect variables.

Ideally we would know the answer to the direct prediction question in each respondent's mind, but we do not. That answer, however, should be based on the sort of information about current and past conditions that the indirect housing questions and indirect questions elicit. So from these questions we are trying to formulate what each respondent's answer to the direct prediction question should be, based on her conditions and history.

3.5 Measures of homelessness

We use both cultural and literal definitions of homelessness. Literal homelessness includes sleeping rough, squatting in abandoned buildings, or living in emergency or crisis accommodation. Cultural homelessness is a broader definition which, in addition to those included in literal homelessness, also includes staying with friends; and living in caravans, cabins, mobile homes, boarding or rooming houses, hostels, hotels or motels. Literal homelessness is close to the definition of homelessness used in the United States, while cultural homelessness is close to the definition used in Australia.

4 Results

4.1 Does private information make predictions about homeless entries better?

Table 2 summarizes the results for the entry equations for three months ahead for the two

types of homelessness (a full set of Probit results can be found in Appendix B, Table A2).

The results show that we can easily reject the null hypothesis that respondents have no relevant private information. Even holding a rich array of public information constant, the private information variables together have significant effects on the probability of becoming homeless. Respondents have useful information that normal intake questionnaires do not pick up.

We see this by examining a measure of goodness-of-fit. For both cultural and literal homelessness, adding private information to public raises the pseudo-R-squared values considerably: from 0.102 to 0.190 for cultural homelessness and from 0.101 to 0.198 for literal homelessness. Private information improves prediction.

To more formally determine the importance of private information we perform a series of Wald tests to determine whether various components of private information help us to predict entries better, given that we already have public information. The resulting χ^2 statistics for each of these tests of joint significance are also presented in Table 2. For the entry equation on cultural homelessness we clearly see that, given public information, we reject the null hypothesis that all coefficients on private information are zero ($\chi^2 = 186.9$). In fact each subset of private information is important; we can reject the null hypothesis that all coefficients on direct housing private information are zero ($\chi^2 = 61.3$) and the null hypothesis that all coefficients on indirect housing private information and non-housing private information are zero ($\chi^2 = 81.5$ and $\chi^2 = 106.8$ respectively). Also, given public information and direct housing private information (not presented in the table), we reject the hypothesis that all coefficients on indirect housing private information and non-housing private information are zero ($\chi^2 = 115.44$).

Wald tests on entry into literal homelessness produce similar results. That is, even when we examine entries into a narrow form of homelessness we can reject the hypothesis that, given public information, all coefficients on private information are zero, whether this is from direct private information, indirect housing information or non-housing private information.

4.2 Displaying improved predictive ability graphically

Receiver-operating-characteristic (ROC) curves are a convenient visual way of summarizing the predictive ability of different equations. For a probit equation, the ROC curve shows the trade-off between the false alarm rate (in our case, the proportion of people falsely predicted to become homeless among all people who do not become homeless) and the hit rate (the proportion of people correctly predicted to be homeless among all people who do become homeless), as the probability threshold required to predict a person becoming homeless varies. (See Shinn et al. 1998 for a famous application of ROC curves to entry into homelessness.)

When should a program operator strive for low false alarm rates, and when should a program operator strive for high hit rates? False alarm rates are socially expensive when either the proportion of the population that does not become homeless is large (so that any false alarm rate corresponds to many false alarms), or when each false alarm is costly, either to the participant, the program or society. High hit rates are desirable if the cost of entering homelessness is great and the prevention program is efficacious—it prevents homelessness at low cost for a large fraction of those who would otherwise be homeless.

Figure 1a shows ROC curves for entering cultural homelessness compiled from various types of information. Interestingly, the private information curve lies above the public information curve: someone trying to predict homelessness who was offered a choice between only private information and only public information would do better by choosing only private information. Such choices are usually not available. The most important result in figure 1a is that the ROC curve for public and private information is considerably above the ROC curve for private information alone. For instance, a program willing to accept a 40 percent false alarm rate would achieve a hit rate of approximately 73 percent with public information alone, but an 84 percent hit rate with both public and private information.

More analytically, we look at the area under the ROC curve. This is a number between 0 and 1; it equals one if the information is perfectly informative, one-half if it is perfectly uninformative, and zero if it is perfectly misleading. Acquiring private information expands the area under the ROC curve from .712 to .808, a difference that is highly significant. Of course, we

have not described how a program operator could acquire the private information, and section 2 has shown that such information cannot be acquired without some sort of cost. Figure 1 shows that attempts to acquire that information may be worthwhile.

Figure 1b is similar to Figure 1a, but for literal homelessness. For a 40 percent false alarm rate, a program could achieve an 80 percent hit rate with public information only, but a 90 percent hit rate with both public and private information.

4.3 Sensitivity analysis

Different versions of private information

The relative usefulness of public and private information for a particular prediction in a particular population depends, of course, on which pieces of information are private and which are public. In this paper, we have identified as public information that which the Australian Specialist Homeless Services Initial Client Form gathers, and that which is available in Centrelink records; everything else we consider private.

Information is partitioned differently in other countries. In this section, we look at how information appears to be partitioned in the United States, and ask whether our results are robust to the US partition.

To see what information homeless program operators gather in the United States, we examined the National Alliance to End Homelessness (NAEH) comprehensive assessment tool (2014). There are two differences with Australia. First, program operators in the United States cannot use information about race or gender in determining eligibility and in assigning people to services. So race and gender are private information in the US, not public. In addition, a person's incarceration record is publicly available in the United States and therefore considered as public information in the United States. As a robustness check, we therefore recalculate our estimates of the public information ROC curve with the US partition rather than the Australian. We concentrate on predicting entry into cultural homelessness within three months. The ROC curve for full information (public plus private) remains the same, of course, since we are merely rearranging the categories in which we place variables.

Second, the NAEH form contains questions about what we have considered private information. Three questions fit into this category: whether the respondent uses drugs weekly, whether he or she engages in risky drinking, and whether he or she experienced violence recently.

The answers to these questions are unverifiable. We do not know whether the answers are consequential—whether they are used to determine who receives valuable services. They may or may not be in different jurisdictions. When the answers are consequential, there is no reason to believe that the answers people give are the same as the answers they would give if the answers were not consequential.

The US public information ROC curve, however, is constructed under the assumption that all people report unverifiable information completely honestly, even when doing so hurts them. In tax compliance and in criminal justice, there is ample evidence that this assumption does not hold. So it is unlikely to hold in homelessness prevention either. A realistic comparison of public information ROC curves, then, must account for some degree of misreporting on these variables.

How much misreporting? We do not know. But tax compliance in the United States provides some basis for comparison. For 2006, the Internal Revenue Service (2012, chart 1) estimated that "amounts subject to substantial information reporting" by third parties but not withholding were misreported by 8 percent (examples include pensions and annuities, social security payments, dividends and interest); "amounts subject to some information reporting" by third parties were misrepresented by 11 percent (examples include partnership/S-corp. income, capital gains, and alimony); and "amounts subject to little or no reporting" were misreported by 56 percent (for example, proprietorship income, farm income, rents and royalties). The last is probably the most relevant category for the areas where the NAEH questionnaire differs from the Australian specialist services questionnaire. We know no reason to expect people at risk of homelessness to be substantially more honest than average Americans, especially when the penalties for misreporting on homelessness prevention intake forms are much less severe than those for misreporting on federal income tax forms.

To account for the possibility of misreporting and to construct ROC curves with

misreporting, we concentrate on the questions about drug use, risky drinking and violence. All of these are binary. We assume that misreporting is independent across respondents: for each respondent the probability of misreporting on question q is the same and independent of whether she misreported on question p≠q. Then if we knew the probabilities of misreporting in both directions, we could construct a modified public information ROC curve that accounts for possible misreporting.[4]

To see how much difference misreporting might make, we experiment with three different scenarios for misreporting. In each scenario for each question we assume a rate of misreporting in each direction. Always the rate of misreporting in the direction that seems most likely to receive services or appear sympathetic is twice the rate of misreporting in the other direction. Table 3 shows the scenarios. Given the IRS data on taxpayer misreporting, these scenarios, even the highest scenario 3 (with corresponding error rates of .54 for drug use, .43 for risky drinking and .53 for violence), are plausible.

The results of these two sets of robustness checks are presented in Table 4 alongside respective pseudo R^2 values. Also, to compare ROC curves quickly, we compute the area under each curve.

As Table 4 shows, the difference between the full information area and Australian public information area is 0.080 (=0.818-0.738). The area under the US public information ROC curve is 0.751; thus using US rather than Australian public information only reduces the gap by about 0.013, or 16 percent of the gap. Table 4 also shows how the area under the modified US public information ROC curve changes under the different misreporting scenarios. As expected, the predictive power of public information decreases as the error rate increases; both the pseudo R^2 and the area under the ROC curve decreases with higher rates of misreporting. Moreover, the area under the ROC curve in scenarios 2 and 3 is not statistically significantly different from the area under an ROC curve that does not use any information about these three variables at all. Intuitively the more misreporting that occurs the less useful the information becomes. Our results suggest that, even under quite

[4] We also looked at correlation in misreporting, but it seemed to make little difference to the results.

reasonable assumptions about misreporting, program operators would be just as accurate in predicting who is likely to enter homelessness if they ignored information on drug and alcohol consumption and experiences of violence as if they included this information.

Our tentative conclusion from this sensitivity analysis is that different plausible partitions of data and different plausible scenarios about misreporting may affect the size of the advantage that using private information can promise, but not greatly.

Longer horizons

Our main results on entries to homelessness concern the three-month-ahead predictions that we discussed above. In this section we look at how well public and private information can forecast homeless entries over longer horizons. If homelessness can be predicted far in advance, program operators might be able to use more time-consuming interventions to prevent homelessness in a targeted fashion

We have no a priori belief about how the different types of information will predict homeless entries in the far future. Therefore we repeated the same exercise we performed for the three-month period for three- and 12-month periods. Journeys Home asks participants directly for their forecast only at a 3-month horizon, and so for longer periods the private information it collects is less relevant. That may bias our results against finding continued importance of private information.

Notice that for each of the longer periods the dependent variable is the probability of at least one homeless entry in the number of months in question following the interview. Thus as the period lengthens, the probability rises (weakly) monotonically. Our results for longer periods, however, are not dominated by entries in the first three months. For instance if we look at entries into cultural homelessness, while almost three-fifths of homeless entries (60.1 per cent) within a six month horizon had already been experienced within the first three months, just over a third (38.5 per cent) of those experienced within a 12 month horizon had been experienced in the first three months.

Figure 2 shows ROC curves for the longer time periods alongside those for three months

presented earlier.[5] The remarkable thing about these curves is that they are quite similar to the curves for three months: the private-information-only curve is always higher than the public-information-only curve, and the heights of these curves are just about the same as the heights of the 3-month curves. However, the private information curve does shift inwards slightly as the time horizon examined increases, closing the gap between the private information and public information curves slightly over a longer time horizon, a result confirmed by an examination of the area under the ROC curves. The area under the ROC curve for private information falls from 0.770 for the three-month horizon to 0.714 for the 12-month horizon. But the area under the public information is virtually the same for the 12-month horizon (0.696) as for the three-month horizon (0.689). See Table 3.

5 Discussion and Applications

How can the predictive private information we have found be used in program design? In this section we suggest a few applications.

Except in rare circumstances, program operators cannot ask applicants for private information. That is why it is private. Asking directly, for instance, "Are you likely to be homeless if we don't help you?" will usually produce no useful information if applicants believe the answer is consequential. Hence, we do not propose asking consequential questions about private information directly. But as the examples of auctions and soup kitchens demonstrate, programs can be designed to allow private information to be revealed, even if no one asks for it.

5.1 Homelessness prevention and program accessibility

Homelessness prevention programs typically rely on participants deciding on their own that they want to participate, and so homelessness prevention programs see only the subset of households who have decided for themselves that they are likely to benefit on net. Entry into homelessness prevention programs is thus a two-step process: first, potential participants decide

[5] Corresponding regression results are presented in Appendix B.

whether they will seek out these programs (at some cost to themselves), and then the programs decide whether to accept the households who seek them out. The first decision can use both private and public information, while the second can generally use only public information.

As the Indonesian conditional cash transfer results (Alatas et al., 2016) suggest, program operators can influence which participants try to enter their programs by altering the accessiblity and publicity of their services, and by how friendly they are. By making the program less than perfectly accessible, program operators are implicitly using participants' private information--that is the alternative to asking for private information on an intake questionnaire.

Think of program operators deciding how hard it is to access their programs. As the program becomes harder to access, some potential participants with lower self-assessed probabilities of becoming homeless will no longer seek services. For any level of difficulty of access, there is an associated minimum probability π of becoming homeless: participants more likely than π to become homeless seek services, and those less likely do not. So for convenience we can think of program operators as choosing π instead of choosing a level of difficulty of access. If accessing homelessness prevention services takes absolutely no effort, everyone will do it (π=0), and the program operator will have to deal with the entire population (or more generally, a sample of the entire population where the probability of selection is independent of the probability of becoming homeless).[6] If accessing homelessness prevention services is harder, only a selected group of potential participants will do it, and that group will be selected on the basis of private information (and public information) indicating that they are likely to enter homelessness soon. See Goodman et al. (2016) for suggestive evidence of a process like this working in New York City.

We can illustrate this process with information from Journeys Home at the three-month horizon. We focus on cultural homelessness. Suppose potential participants use all information (both public and private) to decide whether to seek services. Then for any service-seeking threshold π we can find the households who will seek services. Using only public information, we then estimate a probit equation for the probability of entering cultural homelessness with this sample, as

[6] Remember also that the entire JH sample is disadvantaged.

a program operator would do. Entry into the program thus requires passing two screens: the potential participant herself must judge that she is likely enough to become homeless to apply to the program, and the program operator must confirm that judgment.

To assess how well this two-step process allows program operators to target services, we construct what we call "two-step ROC curves." The base for these curves is the entire population, even though only people who apply to the program stand a chance of being served. The hit rate is the number of people who both apply and are accepted and would have become homeless, divided by the total number of people who would have become homeless. The false alarm rate is the number of people who both apply and are accepted and would not have become homeless, divided by the total number of people who would not have become homeless. As with the standard ROC curves we have seen already, these curves start at the origin—if the program operator serves no one, both the hit rate and false alarm rate are zero. But two-step ROC curves do not usually include (1,1): if the program operator serves everyone who applies, the hit rate will still be less than one because some people who will become homeless will not show up (because their ex ante probability of becoming homeless was low).

A program operator who was intent on a high hit rate and who had no concern about the false alarm rate would not choose to discourage low-probability households like this, because the two-step ROC will be below the simple public information ROC when false alarm rates are very high. But the two-step ROC will be above the simple public information ROC for low false alarm rates. We do not have a dominance relationship. To illustrate, we show the two-step ROCs with thresholds (π) set at levels such that approximately 75 percent, 50 percent and 25 percent of the sample are self-selected service-seekers for both cultural and literal homelessness.

Figure 3 shows two-step ROCs for π=.015, .04, and .09, as well as the simple public information ROC (π=0) for cultural homelessness. As expected, greater values of π raise the two-step ROC for low false alarm rates, but lower it for high false alarm rates; the (two-step) hit rate is about 0.67 for π=.09, compared with 0.50 when only public information is used (π=0). However, the two-step hit rate never rises above 0.70 On the other hand, even if program administrators

provide services to everyone who self-selected to seek services, the false alarm rate will not exceed .22.

Recall that whether a program operator should strive for low false alarm rates or for high hit rates depends on several factors. Low false alarm rates are more important than high hit rates when the proportion of the population who becomes homeless is small and when serving someone who should not be served is socially costly relative to missing someone who should be served.

We do not have information about costs, either social or private, and they will surely vary from program to program. Thus we cannot argue that all program operators should concentrate on a particular part of the ROC curve. But we know something about rates.

Although a large number of variables from Centrelink administrative records (considered as public information) were used to select the Journeys Home population, homeless entries are fairly rare in any three-month period. About 7.5 percent of people who are originally not culturally homeless become culturally homeless in the average three-month period; the comparable figure for literal homelessness is 2.6 percent. Thus false alarms for cultural homelessness are 11.6 times ([100-7.9]/7.9) as common as hits; and for literal homelessness false alarms are 37.5 times ([100-2.6]/2.6) as common as hits. This suggests compulsory programs that select participants based only on public information may not be the most efficient way to use resources.

Probably the best estimate of an efficacy rate for a homelessness prevention program comes from the randomized control trial in New York City (Rolston et al., 2013). The trial was to evaluate the Homebase Community Prevention program where treatment group participants were assigned a case manager, who acted quickly to help the family preserve their housing and develop a long-term plan for housing stability. Services included family and landlord mediation, legal assistance, short-term financial assistance, mental health and substance abuse services, child care, and job search assistance. In that experiment, 8.0 percent of the treatment group entered a shelter within 27-months of randomization, as compared with 14.5 percent of the control group (as we have already noted in the introduction). So this particular program at this particular time was 45 percent efficacious.

Thus it is plausible that many (though not all) program operators in prevention programs

would want to strive for low false alarm rates, and self-selection by private information would help them target services. Of course, the relative costs and benefits of false alarm rates and hit rates might be different in different venues. For instance, homeless health care projects might have relatively low costs for high false alarm rates, since these programs only draw from the pool of people already homeless and since a large portion of homeless people have serious health problems, and relatively high costs for low hit rates, since missing someone could be fatal. Thus homeless health care projects might want to strive for great accessibility.

Private information can also be used to split the pool of potential participants. This is especially helpful when the underlying probit equations have significant interactions. For instance, another way that a homelessness prevention program could take advantage of the correlation between alcohol consumption at risky levels and homelessness entry is to split into two programs with different rules and procedures. One program would offer services to help people recover from drinking and the other would not. The other program parameters would be set to make the programs fairly close in attractiveness, and risky drinkers would seek the program offering the alcohol-recovery service and others would not. Private information would be revealed. Two probits could be fit with public information, one for each program, and different eligibility and acceptance rules could be used (although not so different to make one program or the other too attractive relative to the other). Since it would be feasible for the two programs to re-create the exact workings of the single program, but they are not constrained to do so, such differentiation will generically move the ROC curve out. This can be seen for our risky drinking example in Figure 4.

5.2 Rental insurance programs

The use of private information about homeless entry is not confined to traditional homelessness prevention programs. Consider rental insurance programs, which have been discussed sporadically in the United States (see, e.g., Ellen and O'Flaherty 2010). They are sometimes called "risk mitigation funds," and several jurisdictions in the US operate them on a small scale (US Interagency Council on Homelessness 2016). These programs step in and pay rent when a tenant is unable to do so under specific circumstances. Since we have shown that people

have useful private information about their prospects for becoming homeless, a voluntary tenant-based rental insurance program would be subject to adverse selection: if premiums were actuarially fair based on public information about the population of eligible renters, these programs would disproportionately attract people with private information that they would likely become homeless. Thus the programs would lose money. Our findings imply that rental insurance programs should be designed to counteract adverse selection (for instance, by enrolling through landlords, or mandatorily enrolling all members of some class). In fact, the risk mitigation funds that operate in the US enroll landlords, not tenants.

5.3 The Veterans Administration Universal Screener

Since 2012 the Veterans Administration (VA) in the United States has been using a "universal screener" to identify veterans who are homeless and who are at risk of homelessness. In doing so, it has produced a two-step process like the one we have described in section 6.1, except that the VA assists individuals in the first stage.

The screener has only two questions, and both ask for information we would consider private because it is unverifiable. The first question tries to identify people who are currently homeless, and so only the second, designed to identify people at risk of homelessness, concerns us. This question is: "Are you worried or concerned that in the next two months you may NOT have stable housing that you own, rent, or stay in as part of a household?". This question is much closer to the ideal prediction question that was missing from JH. Veterans who say yes are asked about where they have been living for most of the previous two months, and whether they want to see a social worker or homeless service worker (Montgomery 2014).

These questions are asked of almost all veterans receiving outpatient health care from the Veterans Health Administration—about four million people a year (hence a low false alarm rate is desirable). About 1 percent of the people asked indicate that they are at risk of homelessness (Montgomery 2014), and only about 3 percent of this group appear to become homeless during the next six months (Fargo et al., 2014).

The screener thus asks veterans to use private information to decide whether they want to

begin to access VA homelessness prevention services. The decisions they make are informative: the proportion of veterans who become homeless in a six-month period is far less than the 3 percent observed in this group. The question is also inconsequential: veterans who say they are at risk do not have to accept services, and veterans who do not say they are at risk can still access these services. The "reward" for saying yes must be very small; otherwise more than 1 percent of veterans would say yes. For many veterans, the "reward" may in fact be negative—as in section 5.1—intrusive dealings with bureaucrats, for instance—and so only those with a relatively high probability of becoming homeless say yes.

It is only because this question is inconsequential that it works. It could not be incorporated into a normal intake or assessment form, and the VA does not use it for that purpose. Most program operators could not and would not use it for that purpose because as soon as the information is used to assess program suitability, the program operator cannot assume the private information is accurate.

Thus by creating a two-step process and inducing veterans to reveal some private information, the VA has created a good way of sorting through a large number of potentially homeless people.

Notice that the VA did not have to fit ROC curves in order to develop its universal screener. Nor did it have to prove that the screener it was using was better than any other possible screener. It thought about how people would answer questions (not necessarily truthfully) and what their answers would reveal; then it experimented with different screeners and studied the results; and finally it rolled out a screener that made a major improvement to its outreach and intake process, although it probably was not perfect.

The lesson for program operators thinking about homelessness prevention is that they do not have to fit ROC curves either, or prove that the level of accessibility that they choose is optimal. They should experiment with different levels of accessibility and advertising strategies, try to compare results, and make improvements over time. By analogy, tennis is governed by the laws of physics and it helps to know them; but good tennis players experiment and practice—they do not

have to solve differential equations.

6 Conclusion

We have shown that homeless people and people at risk of homelessness have valuable private information about what is happening to them and what is likely to happen next. Programs that recognize and work with participants' private information can perform better, and those that ignore it run the risk of serious problems.

We have not shown yet that people act on their private information—only that they have it and could act on it. The examples in section 2, however, show that people often do act on their private information, including people experiencing homelessness. The example of the VA universal screener shows the same thing.

The key input into homelessness prevention seems to be the people who seek services, and we know little about how they come to do so. The most proficient service delivery team in the world cannot prevent the homelessness of someone who would not otherwise be homeless. Ethnographic research would probably be helpful in learning how people decide to seek services, and so would more RCTs like the Indonesian study of conditional cash transfers (Alatas et al., 2016).

On a program basis, better advertising and counselling may be helpful in preventing homelessness. If people know more clearly what the symptoms of imminent homelessness are, then they can use their private information about themselves to counter the symptoms and be more effective in deciding when to seek services—and when not to. Better knowledge of the symptoms of imminent homelessness would make those who were more likely to become homeless more likely to seek services, at any level of accessibility, and those who were less likely to become homeless less likely to seek services. At any level of accessibility, the two-stage ROC curve would move out. Unfortunately the data sets on which to base such advertising and counselling are unavailable for the US. But fortunately they are available for Australia.

References

Akerlof, George, 1970,The market for "lemons": Quality uncertainty and the market mechanism, Quarterly Journal of Economics 84(3): 488-500.

Alatas, V., Bannerjee, A., Hanna, R., Olken, B., Purnamasari, R., & Wai-Poi, M. (2016). Self-targeting: Evidence from a field experiment in Indonesia. Journal of Political Economy, 124(2), 371–428.

Burt, Martha R., Carol L. Pearson, and Ann Elizabeth Montgomery, 2005, Strategies for preventing homelessness, U.S. Department of Housing and Urban Development, Office of Policy Development and Research.

Ellen, I.G., & O'Flaherty, B. (2010). Introduction. in I.G. Ellen & B. O'Flaherty (Eds), How to House the Homelessness. New York: Russell Sage Foundation Press.

Evans, William N., James X. Sullivan, and Melanie Wallskog, 2016, Do homelessness prevention programs prevent homelessness? Notre Dame University.

Cobb-Clark, D., Herault, N., Scutella, R., Tseng, Y. (2016). A Journey Home: What Drives How Long People Are Homeless? *Journal of Urban Economics*, 91(January), 57–72.

Goodman, S.F., Messeri, P., & O'Flaherty, B. (2016). Homelessness prevention in New York City: On average, it works. Journal of Housing Economics, 31(1), 14–34.

Gorton, Gary and Andrew Metrick, 2012, Securitized banking and the run on repo, Journal of Financial Economics 104(3): 425-451.

Fargo, J., Byrne, T., Montgomery, A.E., & Roberts, C. (2014). Universal screening for homelessness and risk among veterans: Monitoring housing stability and exploring profiles of risk through repeated screening. Research Brief, VA National Center on Homelessness among Veterans. Retrieved 6 January, 2016, from http://www.endveteranhomelessness.org/sites/default/files/research/UniversalScreeningforHomelessness%26RiskAmong%20Veterans.pdf.

Fox, Jeremy T., and Patrick Bajari, 2013, Measuring the efficiency of an FCC spectrum auction, American Economic Journal Microeconomics 5(1):100-146.

Herndon, N. (2015). Knowledge of future job loss and implications for unemployment insurance. Working paper, Harvard University, Department of Economics.

Mirrlees, J. (1971). An exploration in the theory of optimal taxation. Review of Economic Studies, 38(114), 175–208.

Montgomery, A.E. (2014). Using a universal screener to identify veterans experiencing housing instability. Research Brief, VA National Center on Homelessness among Veterans. Retrieved 6 September, 2016, from http://www.va.gov/HOMELESS/Universal_Screener_to_Identify_Veterans_Experiencing_Housing_Instability_2014.pdf.

Nichols, A.L.,& Zeckhauser, R.J. (1982). Targeting transfers through restrictions on recipients. American Economic Review, 72(2), 372–377.

Nichols, D., Smolensky, E., & Tideman, T.N., (1971). Discrimination by waiting time in merit goods. American Economic Review, 61(3), 312–323.

National Alliance to End Homelessness (2014). Comprehensive Assessment Tool. Retrieved 8 January, 2016, from http://www.endhomelessness.org/library/entry/alliance-coordinated-

assessment-tool-set.

O'Flaherty, B. (2009). When should homeless families get subsidized apartments? Journal of Housing Economics, 18(2), 69–80.

Rolston, H., Geyer, J., Locke, G., & Abt Associates (2013). Final Report: Evaluation of the Homebase Community Prevention Program. New York, Department of Homeless Services.

Roth, Alvin E., 1990, Two-sided Matching: A Study in Game-theoretic modelling and analysis. Cambridge, New York: Cambridge University Press.

Shinn, Marybeth, Beth C. Weitzman, Daniela Stojanovic, James R. Knickman, Lucila Jiminez, Lisa Duchon, Susan James, David H. Krantz, 1998, Predictors of homelessness among families from New York City: From shelter request to housing stability, American Journal of Public Health 88(11): 1651-1657.

Shinn, Marybeth, Jim Baumohl, and Kim Hopper, 2001, The prevention of homelessness revisited, Analysis of Social Issues and Public Policy 1(1): 95-127.

Shinn, M., Greer, A.L., Bainbridge, J., Kwon, J. & Zuiderveen, S. (2013). Efficient targeting of homelessness prevention services for families. American Journal of Public Health, 103(S2): S324–S330.

U.S. Department of Housing and Urban Development (2015). Office of Policy Development and Research, Summary Report: Assessment Tools for Allocating Homelessness Assistance: State of the Evidence.

U. S. Interagency Council on Homelessness (2016). Engaging and supporting landlords through risk mitigation funds: Community profiles, April, Retrieved 26 April, 2016, from https://www.usich.gov/resources/uploads/asset_library/Risk_mitigation_funds_community_profiles.pdf

U.S. Internal Revenue Service (2012). The Tax Gap for Tax Year 2006, Overview. Retrieved 6 January, 2015, from https://www.irs.gov/pub/newsroom/overview_tax_gap_2006.pdf.

Vickrey, W.S. (1961). Counterspeculation, auctions, and competitive sealed tenders. Journal of Finance, 16(1), 8–37.

Wooden, M., Bevitt, A., Chigavazira, A., Greer, N., Johnson, G., Killackey, E., Moschion, J., Scutella, R., Tseng, Y. & Watson, N. (2012). Introducing Journeys Home, Australian Economic Review, 45(3), 368–378.

Table 1: Indicators of public and private information in JH

Information sets	Variables
Public	Age, sex, Aboriginal or Torres Strait Islander, couple, have resident children, education, State, country of birth, whether lives in a major urban area, in State care as a child, have a diagnosed mental health condition, have a long term health condition/disability, in rental or mortgage arrears, housing tenure, moved 6 times or more in previous 6 months, labour force status, Centrelink payment type
Private – direct housing	Reported not being able to stay in current accommodation for next 3 months; Self-assessed housing situation (secure in accommodation, at-risk of homelessness or homeless)
Private – indirect housing	Housing is overcrowded, current place feels safe, current place has adequate sleeping space, current place is in good condition, obstacle to finding another place, satisfaction with housing situation, report having a history of long-term homelessness, history of sleeping rough
Private – non-housing	Total debt; psychological distress; have chronic physical health conditions; experiences of violence; ever in detention (Juvenile justice, prison or remand); drank at risky levels; used marijuana or illicit drugs regularly; due to a shortage of money had to go without food when hungry/had to pawn or sell something/asked a welfare agency for food, clothes, accommodation or money/ asked for financial help from friends or family/could not go out with friends because you could not pay your way/ could not pay electricity, gas or telephone bills on time; no friends; no family contact; most friends work full-time

Table 2: Summary of Probit results: goodness of fit and Wald statistics of joint significance

	Cultural homelessness		Literal homelessness	
Equation (variables included)	Pseudo R^2	Wald test statistics (joint significance of private info variables)[AB]	Pseudo R^2	Wald test statistics (joint significance of private info variables)[AB]
All information [C]	0.190	$\chi^2(34) = 186.9$	0.198	$\chi^2(34) = 130.4$
Public information only [C]	0.102	NA	0.101	NA
All private information only [C]	0.135	NA	0.160	NA
Public + direct housing private	0.133	$\chi^2(3) = 61.3$	0.134	$\chi^2(3) = 38.1$
Public + indirect housing private	0.143	$\chi^2(8) = 81.5$	0.157	$\chi^2(8) = 66.6$
Public + non-housing private	0.151	$\chi^2(23) = 106.8$	0.158	$\chi^2(23) = 64.3$
Public + direct and indirect housing private	0.153	$\chi^2(11) = 108.3$	0.164	$\chi^2(11) = 76.7$

Notes: A. Wald test statistics shows the χ^2 statistics for the tests of joint significance of non-public information variables conditional on public information. $\chi 2$ (#): denotes the $\chi 2$ statistic with # of degrees of freedom

B. All tests have a P value (null: $\chi^2=0$) of <0.001

C. Estimation results for the All information, Public information only and Private information only equations are in Appendix Table A2. Other results are available upon request.

Table 3: Misreporting scenarios

	Drug use		Risky drinking		Recent violence	
	Yes when no	No when yes	Yes when no	No when yes	Yes when no	No when yes
Scenario 1	0.1	0.2	0.1	0.2	0.2	0.1
Scenario 2	0.2	0.4	0.2	0.4	0.4	0.2
Scenario 3	0.3	0.6	0.3	0.6	0.6	0.3

Table 4: Sensitivity of predicting homeless entry with different information sets

			P values: test of significant differences between ROC areas	
	Pseudo R-squared	ROC area	compared with US public ROC	compared with US public ROC without drug/alc/violence info
Public + private	0.190	0.817	<0.001	<0.001
AUS public	0.102	0.738	0.077	0.128
US public	0.114	0.751	-	0.008
US public with errors (scenario 1)	0.107	0.742	0.099	0.066
US public with errors (scenario 2)	0.102	0.737	0.038	0.180
US public with errors (scenario 3)	0.099	0.733	0.007	0.743
US public ROC without drug/alc/violence info	0.099	0.733	0.008	-

Figure 1 Predicting entries to homelessness: ROC curves

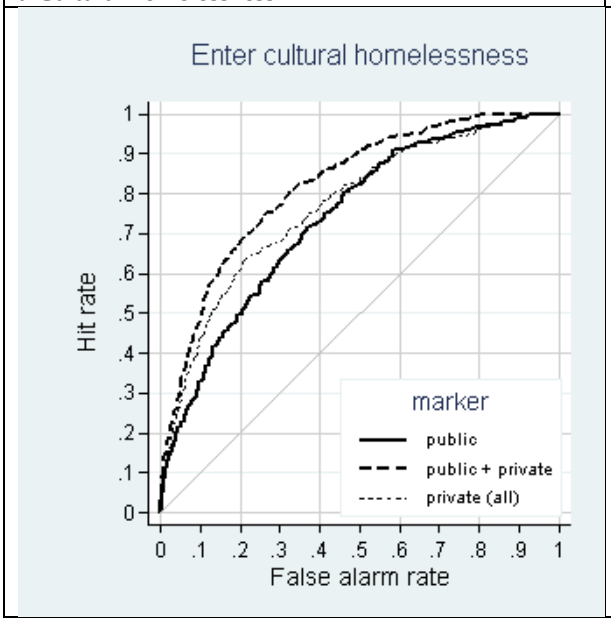
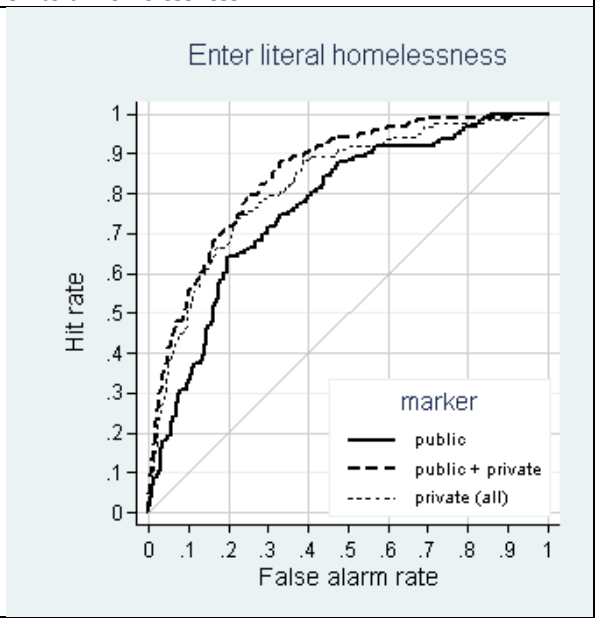
a Cultural homelessness	b literal homelessness
	

Figure 2: ROC curves for entering cultural homelessness within:

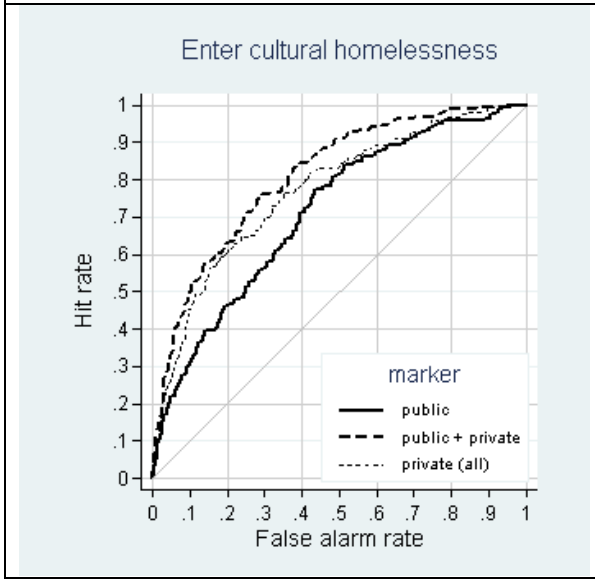
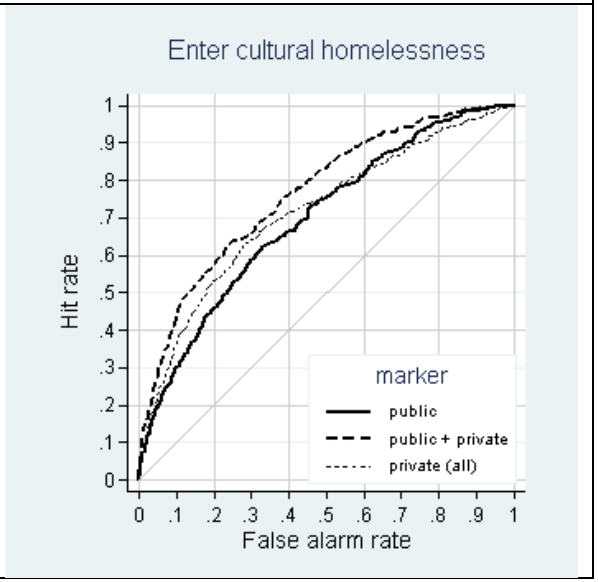
3 months	6 months	12 months
		

Figure 3 Predicting entries to cultural homelessness with self-selection: ROC curves

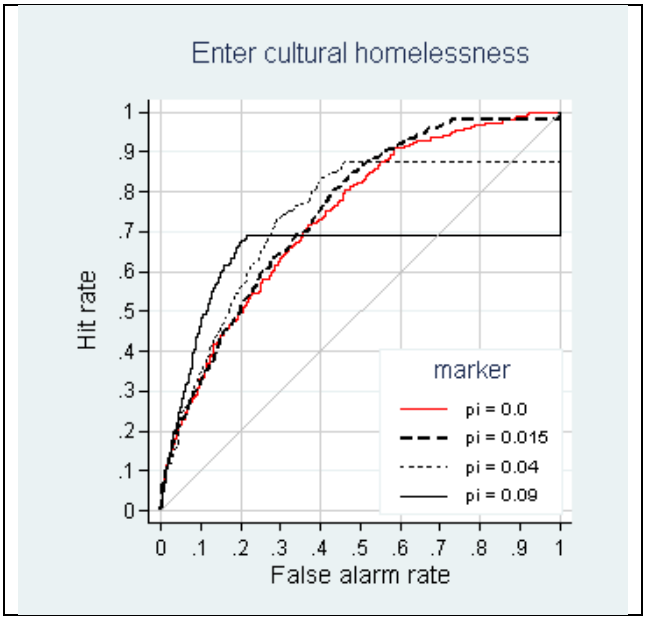
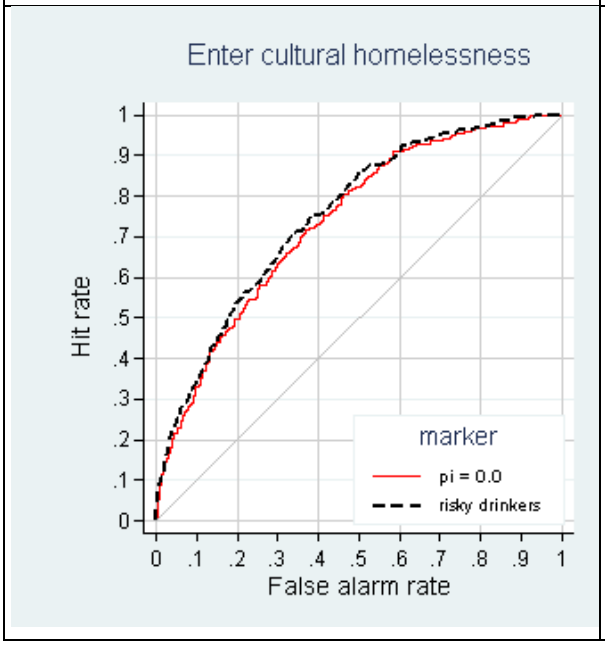
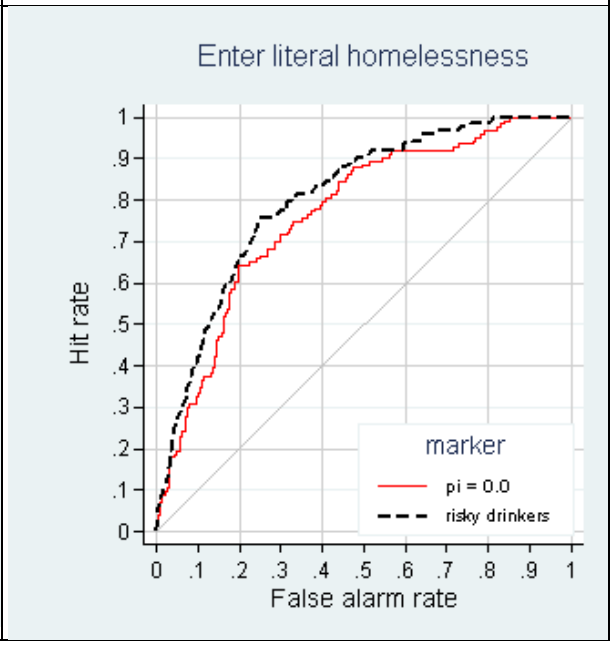


Figure 4 Predicting entries to homelessness with self-selection for risky drinking: ROC curves

a. Cultural homelessness



b. Literal homelessness



Note: For cultural homelessness, rho=.6 represent approximately 90% of sample self-selected to seek services and rho=.72 and .84 represent 75% and 50% respectively.
For literal homelessness rho=.3 represent approximately 90% of sample self-selected to seek services and rho=.45 and .6 represent 80% and 70% respectively. Due to small sample size of literal homelessness observations, a larger proportion of service seekers are needed for second stage equations.

Appendix A: Definitions of variables and sample selection

A.1 Defining homeless entry and exit

The Journeys Home survey collects detailed housing information from individuals at each of their interviews. In addition, from wave 2 onwards, respondents are asked about the timing of all of their moves (recorded as beginning, middle, or end of a month) and the types of accommodation they moved into in between interviews. Homeless status and transitions examined in this paper utilise this housing calendar information. Respondents are then considered as culturally homeless if they are staying with friends; living in a caravan, cabin, mobile home, boarding or rooming house, hostel, hotel or motel; squatting in an abandoned building, living in emergency or crisis accommodation; or sleeping rough (on the streets). Only those who live in a home of their own or their relatives (including parents) are considered as housed. For literal homelessness, those who were squatting in an abandoned building, living in emergency or crisis accommodation; or sleeping rough are considered as homeless and the rest are considered as housed. Individuals who spend time in prison or other institutions are considered as having the same homeless status as that immediately prior to their incarceration.

When predicting entries to homelessness the sample is restricted to person-wave observations where individuals were not homeless at that particular interview whereas when homeless exits are examined the sample is restricted to those homeless at interview. Homeless entry is then defined as a binary variable, where a value of 1 denotes an individual who was housed at their interview but is observed to enter homelessness at least once over the subsequent time period examined (typically 3 months) or 0 where respondents were housed over the entire period.

A.2 Sample selection

Our main analysis sample includes those observations where we observe the individual's current housing status and their housing status over the subsequent three month period (which is obtained from their housing calendar data from a subsequent interview).

In addition, all wave 2 observations were omitted because only a subset of individuals was asked whether they expected to move within 3 months in that wave.

We further excluded observations with missing information in our independent variables (approximately 14% of observations).

This leave us with a final sample of 3828 observations for entering cultural homelessness within 3 months. For literal homelessness, the final sample size is 4663 for homeless entry. Definitions of our independent variables and summary statistics over our resulting main estimation samples are presented in Table A1.

When we examine homeless entries over longer time horizons, we restrict our sample to those observations with valid information for all three time horizons; entering homeless within 3 months, 6 months and 12 months. Thus all wave 5 and 6 observations were excluded as we do not observe individuals' housing transitions over the twelve month horizon. As most individuals are not interviewed at exactly 6 monthly intervals, all wave 4 observations were also omitted as only around half of these observations provided sufficient information on individuals housing transitions over a full twelve month window. As harder to reach individuals will tend to be interviewed later on in the fieldwork period, homeless status may be correlated with the timing of their interviews. Therefore all wave 4 observations were omitted to eliminate the possibility of selection bias from including those observed over the full prospective 12 month window. Thus, the resulting sample utilised when comparing patterns over varying time horizons comparison contains 1919 observations.

Table A1: Variable definitions and summary statistics

	Cultural		Literal	
	Mean	SD	Mean	SD
Entering homelessness within 3 months	0.074	0.263	0.026	0.160
Age groups				
15-24 years	0.400	0.490	0.382	0.486
25-44 years	0.414	0.493	0.409	0.492
45 years plus	0.187	0.390	0.210	0.407
Females	0.509	0.500	0.470	0.499
Aboriginal or Torres Strait Islander	0.179	0.384	0.171	0.377
Couple	0.251	0.433	0.231	0.421
Have resident children	0.301	0.459	0.263	0.440
highest qualification completed				
Tertiary Education	0.329	0.470	0.324	0.468
Completed secondary school	0.118	0.323	0.117	0.321
Yr 10 or 11	0.391	0.488	0.395	0.489
Yr 9 or below	0.154	0.361	0.158	0.365
Education level could not be determined	0.007	0.082	0.006	0.079
State of residence				
New South Wales, Tasmania or ACT	0.295	0.456	0.295	0.456
Victoria	0.206	0.405	0.204	0.403
Queensland	0.274	0.446	0.275	0.447
South Australia	0.068	0.253	0.071	0.256
Western Australia	0.110	0.313	0.109	0.312
Northern Territory	0.046	0.209	0.046	0.210
Country of birth				
Australian born	0.880	0.325	0.876	0.330
Born in main English speaking country	0.062	0.241	0.063	0.243
Born in non English speaking country	0.058	0.234	0.061	0.239
Live outside a major urban area	0.225	0.418	0.219	0.414
Have been in State care as a child	0.263	0.440	0.258	0.437
Have a diagnosed mental health condition	0.677	0.467	0.685	0.465
Have a long term health condition/disability (have last or expected to last longer than 6 months	0.447	0.497	0.454	0.498
In rental or mortgage arrears	0.103	0.304	0.104	0.305
Housing tenure				
Public housing, life tenure or home owner	0.189	0.391	0.160	0.366
Private rental	0.393	0.488	0.385	0.487
Other type of rental	0.419	0.493	0.455	0.498
Moved 6 times or more in previous 6 months	0.024	0.154	0.027	0.162
Labor force status				
employed	0.255	0.436	0.247	0.431
Unemployed	0.254	0.435	0.262	0.440
Not in labour force	0.491	0.500	0.492	0.500
Centrelink payment type				
Not on benefit	0.125	0.331	0.119	0.324
New Start Allowance	0.282	0.450	0.303	0.460
Youth Allowance	0.165	0.371	0.160	0.367
Disability Support Pension	0.235	0.424	0.246	0.431
Other income support payments	0.193	0.394	0.171	0.377

Reported not being able to stay in current accommodation for next 3 months	0.124	0.330	0.138	0.345
Self-assessed housing situation				
Secure in your accommodation	0.867	0.340	0.845	0.362
At risk of homelessness	0.114	0.318	0.126	0.332
Homeless	0.019	0.137	0.030	0.169
Housing is overcrowded	0.059	0.235	0.072	0.259
Current place feels safe	0.925	0.264	0.923	0.267
Current place has adequate sleeping space	0.966	0.182	0.962	0.192
Current place is in good condition	0.923	0.267	0.909	0.287
Obstacle to finding another place: Any	0.193	0.395	0.217	0.412
Life satisfaction: Your housing situation	6.966	2.877	6.736	2.964
Report having a history of long term homelessness	0.194	0.395	0.214	0.410
History of sleeping rough	0.654	0.476	0.673	0.469
Total debt				
No debt	0.313	0.464	0.310	0.463
Less than $1000	0.225	0.418	0.226	0.418
Between $ 1000 and $5000	0.200	0.400	0.200	0.400
More than $5000	0.205	0.404	0.203	0.402
Total debt missing	0.057	0.232	0.061	0.239
Psychological distress (K6 scores)				
Low to moderate psy. distress (<=12)	0.779	0.415	0.773	0.419
Serious psy. distress (13-18)	0.174	0.379	0.179	0.383
Severe psy. distress (19-24)	0.047	0.212	0.048	0.214
Have chronic physical health conditions	0.656	0.475	0.665	0.472
Did not experience violence	0.294	0.456	0.288	0.453
Experienced violence	0.642	0.479	0.649	0.477
Did not respond to violence questions	0.063	0.244	0.064	0.244
Did not experience violence recently	0.798	0.402	0.791	0.407
Experienced violence recently	0.162	0.369	0.170	0.375
Did not respond to recent violence questions	0.040	0.195	0.039	0.194
Ever in detention (Juvenile justice, prison or remand)	0.309	0.462	0.325	0.468
Drank at risky level (>2 drinks per day)	0.570	0.495	0.578	0.494
Used marijuana or illicit drug regularly (weekly)	0.206	0.404	0.225	0.418
Shortage of money: Had to go without food when you were hungry	0.309	0.462	0.330	0.470
Shortage of money: Had to pawn or sell something	0.300	0.458	0.310	0.463
Shortage of money: Asked a welfare agency for food, clothes, accommodation or money	0.291	0.454	0.304	0.460
Shortage of money: Asked for financial help from friends or family	0.480	0.500	0.486	0.500
Shortage of money: Could not go out with friends because you could not pay your way	0.485	0.500	0.495	0.500
Shortage of money: Could not pay electricity, gas or telephone bills on time	0.343	0.475	0.326	0.469
No friend	0.138	0.345	0.141	0.348
Most friends work full-time	0.351	0.477	0.343	0.475
No family contact	0.099	0.299	0.106	0.308
Numbers of observations	3828		4663	

Appendix B: Probit results of entering homelessness

Table A2 Probit results for entering homelessness within 3 months

	Cultural homeless			Literal homeless		
	All info	Public	Private	All info	Public	Private
Age (reference=15-24 years)						
25-44 years	-0.014	-0.002		0.210	0.257**	
	(0.100)	(0.093)		(0.135)	(0.127)	
45 years plus	-0.098	-0.171		0.337**	0.303**	
	(0.128)	(0.124)		(0.154)	(0.148)	
Females	-0.127	-0.192**		-0.171	-0.236**	
	(0.088)	(0.078)		(0.109)	(0.096)	
Aboriginal or Torres Strait Islander	-0.054	-0.063		0.056	0.081	
	(0.104)	(0.102)		(0.121)	(0.118)	
Couple	-0.176**	-0.175**		-0.023	-0.040	
	(0.086)	(0.082)		(0.113)	(0.108)	
Have resident children	0.064	0.049		-0.090	-0.156	
	(0.106)	(0.100)		(0.140)	(0.138)	
Education (reference = Tertiary)						
Completed secondary school	0.240*	0.154		0.166	0.148	
	(0.128)	(0.120)		(0.148)	(0.145)	
Yr 10 or 11	0.305***	0.322***		-0.013	0.039	
	(0.089)	(0.083)		(0.109)	(0.099)	
Yr 9 or below	0.320***	0.421***		-0.142	-0.029	
	(0.115)	(0.106)		(0.148)	(0.137)	
Education level could not be determined	0.303	0.483		0.096	0.343	
	(0.292)	(0.349)		(0.231)	(0.286)	
State of residence (reference=NSW, Tasmania or ACT)						
Victoria	0.060	0.061		-0.003	-0.004	
	(0.110)	(0.103)		(0.138)	(0.129)	
Queensland	0.238**	0.214**		0.227*	0.148	
	(0.100)	(0.094)		(0.120)	(0.115)	
South Australia	0.050	-0.067		0.135	-0.036	
	(0.150)	(0.145)		(0.194)	(0.179)	
Western Australia	0.548***	0.495***		0.125	0.106	
	(0.114)	(0.110)		(0.161)	(0.148)	
Northern Territory	0.156	0.163		0.207	0.211	
	(0.193)	(0.193)		(0.215)	(0.210)	
Country of birth (reference=Australian born)						
Born in mainly English speaking country	0.003	-0.032		-0.219	-0.263	
	(0.155)	(0.151)		(0.168)	(0.170)	
Born in non-English speaking country	0.133	-0.013		-0.078	-0.151	
	(0.179)	(0.174)		(0.204)	(0.197)	
Live outside a major urban area	0.155*	0.124		-0.249**	-0.309***	
	(0.088)	(0.085)		(0.122)	(0.117)	
In State care as a child	0.114	0.172**		0.196**	0.239**	
	(0.082)	(0.074)		(0.099)	(0.093)	
Have a diagnosed mental health condition	0.066	0.163**		0.008	0.080	
	(0.087)	(0.081)		(0.114)	(0.108)	
Have a long term health condition or disability	0.035	0.112		0.287***	0.342***	
	(0.081)	(0.077)		(0.105)	(0.100)	
In rental or mortgage arrears	0.123	0.257***		0.063	0.161	

	Cultural homeless			Literal homeless		
	All info	Public	Private	All info	Public	Private
	(0.101)	(0.094)		(0.133)	(0.120)	
Housing tenure (reference = Public housing, life tenure, or home owner)						
Private rental	0.564***	0.504***		0.132	0.141	
	(0.122)	(0.118)		(0.154)	(0.143)	
Other type of rental	0.525***	0.587***		0.243	0.347**	
	(0.121)	(0.115)		(0.150)	(0.138)	
Moved 6 times or more in previous 6 months	0.263*	0.748***		0.309*	0.743***	
	(0.157)	(0.144)		(0.172)	(0.160)	
Labour force status (reference = employed)						
Unemployed	0.152	0.246**		0.208	0.280*	
	(0.113)	(0.108)		(0.156)	(0.152)	
Not in labour force	0.168	0.205*		0.112	0.212	
	(0.122)	(0.113)		(0.158)	(0.153)	
Centrelink payment type (reference=no on benefit)						
New Start Allowance	-0.127	0.030		-0.181	-0.004	
	(0.138)	(0.136)		(0.200)	(0.196)	
Youth Allowance	0.121	0.158		-0.032	-0.020	
	(0.149)	(0.143)		(0.223)	(0.213)	
Disability Support Pension	-0.349**	-0.252		-0.189	-0.127	
	(0.166)	(0.160)		(0.214)	(0.216)	
Other payments	-0.441**	-0.312*		-0.140	-0.037	
	(0.188)	(0.177)		(0.253)	(0.250)	
Reported not being able to stay in current accommodation for next 3 months	0.293***		0.364***	0.085		0.078
	(0.100)		(0.098)	(0.127)		(0.126)
Self-assessed housing situation (reference = secure in your accommodation)						
Assess that they are at risk of homelessness	0.073		0.140	-0.002		0.025
	(0.116)		(0.111)	(0.129)		(0.122)
Assess that they are homeless	0.461**		0.550***	0.359		0.460**
	(0.218)		(0.207)	(0.220)		(0.205)
Housing is overcrowded	0.142		0.145	0.251*		0.282**
	(0.142)		(0.136)	(0.145)		(0.142)
Current place feels safe	0.375***		0.480***	0.053		0.068
	(0.135)		(0.129)	(0.139)		(0.137)
Current place has adequate sleeping space	-0.097		-0.111	0.187		0.202
	(0.190)		(0.182)	(0.213)		(0.205)
Current place is in good condition	-0.023		-0.021	-0.084		-0.065
	(0.133)		(0.128)	(0.134)		(0.134)
Obstacle to finding another place: Any	0.233**		0.295***	0.155		0.145
	(0.093)		(0.087)	(0.111)		(0.108)
Life satisfaction: Your housing situation	-0.029*		-0.020	-0.045**		-0.050***
	(0.015)		(0.015)	(0.019)		(0.018)
Report having a history of long term homelessness	0.246***		0.155*	0.119		0.166*
	(0.094)		(0.088)	(0.100)		(0.099)
History of sleeping rough	0.055		0.036	0.294**		0.323**
	(0.088)		(0.085)	(0.135)		(0.136)
Total debt (reference= no debt)						
less than $1000	-0.049		-0.060	-0.048		-0.030
	(0.102)		(0.100)	(0.135)		(0.135)
between $ 1000 and $5000	0.167		0.102	0.068		0.065

	Cultural homeless			Literal homeless		
	All info	Public	Private	All info	Public	Private
	(0.107)		(0.103)	(0.143)		(0.139)
more than $5000	-0.056		-0.058	0.005		0.043
	(0.107)		(0.102)	(0.144)		(0.137)
Total debt missing	0.079		0.044	0.286*		0.308*
	(0.146)		(0.144)	(0.163)		(0.161)
Psychological distress (reference= low distress)						
Moderate distress	-0.089		-0.063	-0.053		-0.002
	(0.098)		(0.093)	(0.108)		(0.105)
Severe distress	-0.413**		-0.358**	-0.015		0.057
	(0.163)		(0.155)	(0.180)		(0.181)
Have chronic physical health conditions	0.007		-0.036	-0.168*		-0.094
	(0.080)		(0.075)	(0.101)		(0.097)
Violence (reference=Did not experience violence (physical or sexual) as a child)						
Experienced violence as a child (physical or sexual)	0.190*		0.218**	-0.033		-0.002
	(0.100)		(0.091)	(0.111)		(0.107)
Did not respond to violence questions	0.169		0.202	-0.622*		-0.677*
	(0.195)		(0.189)	(0.372)		(0.350)
Experienced violence in the last 6m	0.184**		0.177**	-0.036		-0.048
	(0.089)		(0.086)	(0.109)		(0.110)
Did not respond to recent violence questions	0.031		-0.024	0.199		0.130
	(0.198)		(0.191)	(0.271)		(0.259)
Ever in detention	-0.145		-0.135*	-0.095		-0.009
	(0.092)		(0.080)	(0.108)		(0.093)
Drank at risky levels	0.183**		0.290***	0.137		0.143
	(0.078)		(0.071)	(0.096)		(0.093)

	Cultural homeless			Literal homeless		
	All info	Public	Private	All info	Public	Private
used drug regularly (weekly)	0.043		0.120	0.153		0.174*
	(0.086)		(0.082)	(0.097)		(0.093)
Shortage of money: Had to go without food when you were hungry	0.138		0.185**	0.296***		0.301***
	(0.086)		(0.082)	(0.114)		(0.113)
Shortage of money: Had to pawn or sell something	0.233***		0.261***	0.112		0.091
	(0.085)		(0.080)	(0.105)		(0.103)
Shortage of money: Asked a welfare agency for food, clothes, accommodation or money	0.223***		0.157**	0.127		0.150
	(0.079)		(0.076)	(0.097)		(0.100)
Shortage of money: Asked for financial help from friends or family	0.027		0.038	-0.042		-0.053
	(0.080)		(0.078)	(0.106)		(0.103)
Shortage of money: Could not go out with friends because you could not pay your way	-0.010		-0.027	0.033		0.036
	(0.085)		(0.080)	(0.107)		(0.106)
Shortage of money: Could not pay electricity, gas or telephone bills on time	-0.249***		-0.272***	-0.191*		-0.205**
	(0.081)		(0.076)	(0.106)		(0.100)
No friend	0.075		0.023	-0.046		-0.014
	(0.101)		(0.098)	(0.111)		(0.108)
Most friends work full-time	0.009		-0.007	-0.138		-0.101
	(0.077)		(0.075)	(0.101)		(0.099)
No family contact	0.273***		0.212**	-0.020		-0.020
	(0.101)		(0.098)	(0.126)		(0.124)

	Cultural homeless			Literal homeless		
	All info	Public	Private	All info	Public	Private
Constant	-3.059***	-2.557***	-2.372***	-2.824***	-2.782***	-2.501***
	(0.344)	(0.187)	(0.269)	(0.369)	(0.231)	(0.271)
Pseudo R-squared	0.190	0.102	0.135	0.198	0.101	0.160
Numbers of observations	3,828	3,828	3,828	4,663	4,663	4,663

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table A3: Predicting homeless entry, predictive power over various time horizons

	Homelessness entry prediction		
	Public	All private	All info
		Within 3 months	
Pseudo R squared	0.084	0.135	0.179
Area under ROC curve	0.712	0.770	0.808
Std error of ROC curve area	0.020	0.019	0.017
Differences: ROC area - public info ROC area	-	0.058	0.096
P value: difference=0	-	0.018	<0.001
		Within 6 months	
Pseudo R squared	0.076	0.118	0.160
Area under ROC curve	0.689	0.745	0.780
Std error of ROC curve area	0.017	0.016	0.014
Differences: ROC area - public info ROC area	-	0.056	0.092
P value: difference=0	-	0.006	<0.001
		Within 12 months	
Pseudo R squared	0.089	0.105	0.156
Area under ROC curve	0.696	0.714	0.763
Std error of ROC curve area	0.014	0.014	0.013
Differences: ROC area - public info ROC area	-	0.018	0.066
P value: difference=0	-	0.308	<0.001
Number of observations		1918	

Note: full regression results are available from the authors upon request.