Appendix A Model Specification

The Markovian model is specified as follows:

For each individual, $i = 1, ..., N$, define B_{it} to be a binary variable summarising whether or not an individual's annual household income is below the threshold in time t. B_{it} is equal to one if an individual's household income in time t is below the threshold, and zero otherwise. Define B_{it-1} to be a binary indicator equal to one if an individual's household income is below the threshold in time $t-1$, and zero otherwise. Let R_{it} be a binary variable for panel retention, taking the value one if an individual has household income information in time t and zero if the individual is only observed in time $t-1$. That is, B_{it} is only observed when $R_{it} = 1$. For each time period, individuals can be characterised by the latent propensity of household income falling below the threshold, b^*_{it}, latent propensity of initial household income being below the threshold, b^*_{it-1}, and latent propensity of sample retention, r^*_{it}, that take the form:

$$b^*_{it} = [(B_{it-1})\gamma_1' + (1 - B_{it-1})\gamma_2']\boldsymbol{x}_{it-1} + u_{it} \tag{1}$$

$$b^*_{it-1} = \beta'\boldsymbol{z}_{it-1} + v_{it-1} \tag{2}$$

$$r^*_{it} = \psi'\boldsymbol{w}_{it-1} + \varepsilon_{it}, \tag{3}$$

where

$$\begin{aligned} u_{it} &= \mu_i + \delta_{it} \sim N(0,1) \\ v_{it-1} &= o_i + \pi_{it-1} \sim N(0,1) \\ \varepsilon_{it} &= \eta_{we} + \xi_{it} \sim N(0,1), \end{aligned}$$

and

$$\begin{aligned} B_{it} &= I(b^*_{it} > 0) \\ B_{it-1} &= I(b^*_{it-1} > 0) \\ R_{it} &= I(r^*_{it} > 0). \end{aligned}$$

The vectors of covariates, $\boldsymbol{x}_{it-1}$, $\boldsymbol{z}_{it-1}$ and $\boldsymbol{w}_{it-1}$, characterise individual i's household in their base year (assumed to be pre-determined); γ, β, and ψ are parameter vectors and u_{it}, v_{it}, and ε_{it} are the three error terms, assumed to be distributed trivariate standard normal and defined as the sum of an individual specific effect (μ_i, o_i, η_i) and an orthogonal white noise error (δ_{it}, π_{it-1}, ξ_{it}) that follows a normal distribution. The $i(\cdot)$ are binary indicator functions equal to one if the underlying latent propensity exceeds some unobserved value (which can be zero without loss of generality) and equal to zero otherwise. Additionally, equation 3 can be viewed as a selection equation: it governs whether or not individuals contribute to the estimation of the transition probabilities (γ_1 and γ_2) in equation 1.

The log-likelihood contribution of individual i, whose household income is observed in the initial period, is given by the following sample log-likelihood function:

$$
\begin{aligned}
logL_i =& B_{it-1}R_{it}log[\Phi_3(k_i\gamma_1'\boldsymbol{x}_{it-1}, m_i\psi'\boldsymbol{w}_{it-1}, q_i\beta'\boldsymbol{z}_{it-1}; k_im_i\rho_{31}, k_iq_i\rho_{21}; m_iq_i\rho_{32})] \\
&+ (1-B_{it-1})R_{it}log[\Phi_3(k_i\gamma_2'\boldsymbol{x}_{it-1}, m_i\psi'\boldsymbol{w}_{it-1}, q_i\beta'\boldsymbol{z}_{it-1}; k_im_i\rho_{31}, k_iq_i\rho_{21}; m_iq_i\rho_{32})] \\
&+ (1-R_{it})log[\Phi_2(m_i\psi'\boldsymbol{w}_{it-1}, q_i\beta'\boldsymbol{z}_{it-1}; m_iq_i\rho_{32})],
\end{aligned}
$$

where $k_i \equiv 2B_{it}-1$, $m_i \equiv 2R_{it-1}-1$, and $q_i \equiv 2B_{it-1}-1$. $\Phi_3(\cdot)$ and $\Phi_2(\cdot)$ are the cumulative density functions of the trivariate and bivariate standard normal distributions respectively. The first term of the log-likelihood function corresponds to the contribution of someone whose household income was below the threshold in the base period and who was retained in the sample. The second term is the contribution of an individual whose household income was above the threshold in the initial period and who was retained in the sample. The third term is the contribution of an individual whose household income was observed in the initial period, but whose household income was not observed in the following wave (i.e. the individual and/or their household attrited from the sample). The use of base year characteristics as covariates means that transition probabilities can be predicted for the attritor subsample (individuals with $R_{it}=0$), using estimates that are robust to non-random attrition.

There are three correlations that are estimated to parameterise the unobserved heterogeneity in the model. These are:

$$\rho_{21} \equiv corr(u_{it}, v_{it-1}) = cov(\mu_i, o_i),$$

which gives the association between unobservable characteristics determining whether an individual's household income is below the threshold in time $t-1$ and below the threshold in time t,

$$\rho_{31} \equiv corr(u_{it}, \varepsilon_{it}) = cov(\mu_i, \eta_i),$$

which gives the relationship between unobservable characteristics determining sample retention and the probability of household income being below the threshold in time t, and

$$\rho_{32} \equiv corr(v_{it-1}, \varepsilon_{it}) = cov(o_i, \eta_i),$$

which gives the relationship between unobserved heterogeneity determining sample retention and the probability of an individual's household income being below the threshold in time $t-1$.

A test of no correlation between these cross equation error terms may allow for a simplified model. Other things equal, if $\rho_{21} = \rho_{32} = 0$, there is no initial conditions problem (i.e. initial conditions are exogenous) and having household income below the threshold in time t could be simultaneously modelled with sample retention by a bivariate probit regression. Similarly, if $\rho_{31} = \rho_{32} = 0$, then the process governing panel attrition is exogenous. In this case, household income being below the threshold in time t could be simultaneously modelled with the initial conditions by a bivariate probit regression. Finally, if $\rho_{21} = \rho_{31} = \rho_{32} = 0$, then both initial conditions and retention are exogenous and household income position in relation to the threshold in time t could be estimated by a univariate probit regression.

If $\rho_{21} = \rho_{31} = \rho_{32} = 0$ cannot be rejected, then statistical identification of the model parameters, γ, β, and ψ, requires exclusion restrictions. Specifically, we need to find a set of instrumental variables [IVs] that affect initial state (in time $t-1$) but have no direct effect on household income being above or below the threshold in the following period (time t). Similarly, we need a set of IVs that affect sample retention but have no direct effect on household income being above or below the threshold in time t. For the IVs to be valid, they need to both be relevant and excludable. We test these assumptions in Table 7.

The IVs for the initial conditions equation 2 are categories of father's occupation (including a category for this information being missing; see Table 10). We also include parental education and mother's occupation categories in the regression, but these variables did not satisfy the exclusion restriction. A valid instrument for retention equation 3 is an indicator variable for whether the respondent was classified by their interviewer as very friendly during the interview. We assume that friendliness during the interview has no direct effect on household income being above or below the threshold once its effect on sample retention has been accounted for. We also include a binary variable for whether or not the respondent was a continuing sample member [CSM] or temporary sample member [TSM] (as in Schotte et al., 2018) and a binary variable for attentiveness during the interview, but these variables do not satisfy the exclusion restriction. That the instruments satisfy the relevance condition is shown in Table 7 and discussed below.

Our specification allows for differing effects of covariates on transitions depending on whether an individual's household income was above or below the threshold in the base period. The model implies the following equations for the conditional probability of household income remaining below the threshold (s_{it}) and the conditional probability of household income falling below the threshold (e_{it}), respectively:

$$s_{it} \equiv Pr(B_{it} = 1 | B_{it-1} = 1) = \frac{\Phi_2(\gamma_1' \boldsymbol{x}_{it-1}, \beta' \boldsymbol{z}_{it-1}, \rho_{21})}{\Phi(\beta' \boldsymbol{z}_{it-1})}$$

$$e_{it} \equiv Pr(B_{it} = 1 | B_{it-1} = 0) = \frac{\Phi_2(\gamma_2' \boldsymbol{x}_{it-1}, -\beta' \boldsymbol{z}_{it-1}, -\rho_{21})}{\Phi(-\beta' \boldsymbol{z}_{it-1})}$$

A.1 Model specification tests

We begin by testing the exogeneity of initial conditions equation 2 and sample retention equation 3 with respect to household income being below the threshold in time t. That is, we test the separate and joint significance of the correlation coefficients ρ_{21}, ρ_{31}, and ρ_{32}. We also test the validity of our exclusion restrictions by treating the non-linear functional form of the model as sufficient for identification, and using father's occupation and the indicator variable for friendliness as over-identifying restrictions for equations 2 and 3 respectively. The results of these tests are presented in Table 7.

The evidence in panels (a) and (b) suggests that the estimation strategy for household income transitions around the threshold is appropriate. From panel (a), we observe significant correlation between unobservables affecting initial position in relation to the threshold and falling below the threshold in time t (ρ_{21}), as well as initial position (above or below) in time $t-1$ and sample retention (ρ_{31}). Moreover, panel (b) shows exogeneity of initial conditions and sample retention is rejected at the 1% significance level, as is joint exogeneity. Both initial conditions and sample retention are thus be considered endogenous to this model.

Table 7 panel (c) shows that the indicator variables for father's occupation are significantly correlated with household income being below the threshold in time $t-1$, but are excludable from the main equation. Similarly, being friendly during the interview affects sample retention, but has no association with household income being below the threshold in time t. As such, we are confident that these controls allow for the identification of the system of equations.

Table 7: Estimates of model correlations and model test statistics

	Estimate	P-value
(a) Correlation coefficients between unobservables		
Base-year status and conditional current status (ρ_{21})	-0.47	0.00
Retention and conditional current status (ρ_{31})	0.31	0.00
Retention and base-year status (ρ_{32})	0.13	0.00
Null hypotheses	Test statistic	P-value
(b) Wald test for exogeneity of selection equations		
Exogeneity of initial conditions ($\rho_{21} = \rho_{32} = 0$)	21.51	0.00
Exogeneity of sample retention ($\rho_{31} = \rho_{32} = 0$)	21.48	0.00
Joint exogeneity ($\rho_{21} = \rho_{31} = \rho_{32} = 0$)	42.67	0.00
(c) Instrument validity		
Exclusion of father's occupation from transition eqn (d.f.=10)	17.66	0.13
Exclusion of friendliness indicator from transition eqn (d.f.=2)	0.83	0.66
Exclusion of father's occupation & friendliness indicator from transition eqn (d.f.=12)	17.12	0.25
Relevance of father's occupation in initial conditions eqn (d.f.=5)	26.21	0.00
Relevance of friendliness indicator in retention eqn (d.f.=1)	10.41	0.00

Source: Authors' own calculations using NIDS Waves 1 to 5 pooled sample.
Notes: Simulated pseudo maximum likelihood estimation with 251 random draws. Data are weighted using post-stratification weights from the base period. Standard errors are robust to arbitrary levels of intra-household correlation and the presence of repeated observations on the same individual.

Appendix B Checks and considerations

B.1 Sensitivity to the choice of probability cut-off

We recognise that the size and mean characteristics of our funding classes may be sensitive to the value set as the probability threshold. In the main analysis, we follow Schotte et al. (2018) and López-Calva and Ortiz-Juarez (2014) in setting the cut-off separating the vulnerable and stable missing middle at the observed rate of those with household income above the threshold that fell below in the next period (37.79%). That is, those with predicted probabilities of household income falling below the threshold of more than 37.79% are classified as vulnerable.

In principle, however, it would be possible to make an argument for the use of another probability threshold that sets a different cut-off for the vulnerable. By way of illustration, a stricter (higher) threshold could be implemented as the cut-off if, for example, scarcity of funds is a concern. In this hypothetical case, we are interested in the extent to which this would change who is defined as vulnerable. A concern would be that through the implementation of a stricter threshold, particularly vulnerable individuals will be classified as not requiring support.

Key to understanding the sensitivity of our profiling to these thresholds, is to understand how the estimates of mean household characteristics change when the missing middle is split by different probability thresholds. To explore this, we examine the average differences in key characteristics of the vulnerable and stable missing middle groups when the threshold separating these groups is raised from 37.79% to 50% and 75%, respectively. The vulnerable missing middle will then be defined as those whose predicted probabilities of their household income falling below R350 000 are more than 0.5 and more than 0.75, respectively. Intuitively, by raising the threshold the share of the missing middle classified as vulnerable will be lower, and the share classified as stable will be larger. That is, as the threshold value increases, those whose predicted probabilities of household income between 37.79% and 75% will be reclassified from vulnerable to stable. Table 8 shows these results.

Throughout this section, we consider which individuals in the missing middle are classified as vulnerable and stable when setting different values as the probability threshold. However, the same logic applies to the threshold separating the persistently eligible from the transiently eligible. Table 14 in Appendix C shows the average of all characteristics of the household and household head for the vulnerable and stable missing middle when the probability threshold is set to 0.5 and 0.75, respectively.

When defined by a predicted probability of falling below R350 000 being greater than 0.5, 35% of the missing middle is classified as vulnerable. When the vulnerable group is restricted to those whose predicted probability of falling below R350 000 is above 0.75, only 16% of the missing middle remain classified as vulnerable. We examine the mean characteristics of the vulnerable and stable missing middle groups under these compositional changes. A couple of main points stand out. First, as individuals move from vulnerable to stable groups, it is apparent that although mean characteristics of both groups fall, the stable missing middle do not experience as large a drop in their mean characteristics as the vulnerable. This is most evident in income per capita, where one observes a greater fall in mean per capita income for the vulnerable missing middle group when relatively less vulnerable individuals are pushed into the stable class, compared to the fall in mean per capita income for the stable group when they absorb some individuals moving up from the vulnerable group.

Second, when only 16% of the missing middle remains in the vulnerable group (those whose predicted probability of falling below the threshold is greater than 0.75), their average characteristics very closely reflect those of the transiently eligible group below the threshold. If one considers vulnerability from a mobility perspective, it is thus clear that these individuals in particular are warranting of policy attention. This result further speaks to the potential benefits of a more nuanced or differentiated funding instrument, especially for those in the missing middle. Indeed, it may be more tangible and palatable to extend funding to the most vulnerable 16% of the missing middle, rather than the full spectrum of incomes off which it is currently defined.

Income thresholds are always likely to exist for operational purposes and thus speculating about an increase in a probability threshold is purely hypothetical. Nonetheless, it is reassuring that the above results suggest that by raising the threshold there does not appear to be a considerable risk of mis-identifying highly vulnerable individuals as 'stable' – at least not in ways that substantially change average access to credit markets, assets, and household employment. This could be because the current household income threshold of R350 000 already does a relatively good job of capturing household vulnerability.

Table 8: Average characteristics of households and household heads for those in the vulnerable vs. stable missing middle by probability threshold value

	Threshold = 37.79%		**Threshold = 50%**		**Threshold = 75%**	
	Vulnerable	Stable	Vulnerable	Stable	Vulnerable	Stable
Share of missing middle	47.40%	52.60%	34.52%	65.48%	15.66%	84.34%
(a) Characteristics of the head						
Race						
African	0.76	0.16	0.83	0.24	0.94	0.35
Coloured	0.13	0.13	0.11	0.14	0.03	0.15
Asian/Indian	0.05	0.06	0.04	0.07	0.03	0.06
White	0.06	0.65	0.02	0.56	0.00	0.44
Household head is employed	0.64	0.79	0.65	0.75	0.52	0.75
Female household head	0.48	0.26	0.51	0.29	0.59	0.32
Head has a post-school qualification	0.37	0.71	0.32	0.67	0.15	0.62
(b) Household characteristics						
Household annual income	446 873.20	456 642.00	443 649.80	456 420.10	440 593.50	454 131.30
Per capita monthly income[a]	8 835.43	11 986.31	8 218.88	11 691.68	6 233.05	11 283.60
Share of income from labour market	0.80	0.88	0.79	0.88	0.69	0.88
At least one resident has:						
Financial assets	0.94	0.94	0.93	0.94	0.92	0.94
Pension assets	0.31	0.52	0.29	0.48	0.24	0.45
Home loan/bond	0.24	0.53	0.18	0.51	0.09	0.46
Bank loan	0.34	0.28	0.32	0.31	0.29	0.32
Vehicle finance	0.30	0.35	0.27	0.35	0.16	0.36
Credit card	0.32	0.47	0.31	0.45	0.23	0.43
Household dependency ratio	0.63	0.49	0.64	0.51	0.79	0.51
Household employment rate[b]	0.57	0.74	0.54	0.72	0.48	0.69
Household is in an urban location	0.78	0.95	0.75	0.93	0.66	0.90

Source: Authors' own calculations using NIDS Waves 1 to 5 pooled sample.
Notes: Data are weighted using post-stratification weights from the base period.
[a] Value in December 2017 Rands. Labour market income reflects after tax income. Share from labour market reflects the share of household income from the labour market. [b] Of working age adults.

B.2 The assumption underlying pooled transitions

By pooling waves of observations, we implicitly assume that the economic conditions affecting transitions in each wave remain unchanged throughout the panel. As noted by Zizzamia et al. (2016), this is an important assumption given that changes in the macroeconomic environment will affect the chances of household income rising or falling. In order to examine whether the process generating a rise or fall in household income is reasonably consistent between waves, we fit the multivariate probit regression to each pair of waves independently, full results of which are presented in Table 11 in Appendix C. We do not include interactions between initial state and the covariates in these specifications, since the model does not converge when interactions are included. The probit specification without interactions mirrors that used by López-Calva and Ortiz-Juarez (2014) for identifying a vulnerable middle class.

A majority of covariates remain consistent in terms of sign and significance. In most of the cases where the sign of the effect differs, the effect of the covariate is not statistically different from zero. This ameliorates some concerns about changing environments between waves. Nonetheless, it is the predicted conditional propensities of household income moving above or falling below the threshold that are key to our analysis, rather than the coefficient estimates. Figures 5 and 6 therefore compare the distribution of predicted propensities for the eligible and missing middle groups when predicted from the multivariate probit on the pooled sample and each pair of waves, respectively.

The fact that the distributions from the pooled sample and each pair of waves track each other closely reassures us that we are not misidentifying funding classes by classifying them off predicted conditional propensities using the pooled sample rather than separate wave by wave transitions. However, the distributions in the figures do draw attention to the comparability of predicted propensities between pairs of waves, particularly between the Wave 1-2 transition and the Wave 2-3 transition. For the Wave 2-3 transition in Figure 5, at each income level the average predicted probability of household income moving above the threshold is lower compared to the Wave 1-2 transition. This may owe to the fact that in Wave 1, the sample of respondents is nationally representative, but in each subsequent wave, attrition affects the extent to which the sample (and thus the next 'base period') remains representative. Therefore, those who would be more likely to transition above the threshold between Waves 2 and 3 were more likely to be lost from the sample between Waves 1 and 2. By design of the sample, we will thus be underestimating the size of the transiently eligible group. Although the post-stratification weights aim to ensure that each cross section is representative of the population at the time, expecting the weights to be able to fully account for information lost due to attrition is, perhaps, too ambitious.

For the Wave 2-3 transition in Figure 5, at each income level the average predicted probability of household income moving above the threshold is lower compared to the Wave 1-2 transition. This may owe to the fact that in Wave 1, the sample of respondents is nationally representative, but in each subsequent wave, attrition affects the extent to which the sample (and thus the next 'base period') remains representative. Therefore, those who would be more likely to transition above the threshold between Waves 2 and 3 were more likely to be lost from the sample between Waves 1 and 2. By design of the sample, we could thus be underestimating the size of the transiently eligible group. On the other hand, one could argue that the size of the transiently eligible group is over-estimated. This would be that case if we are disproportionately losing people who would have moved above the threshold.

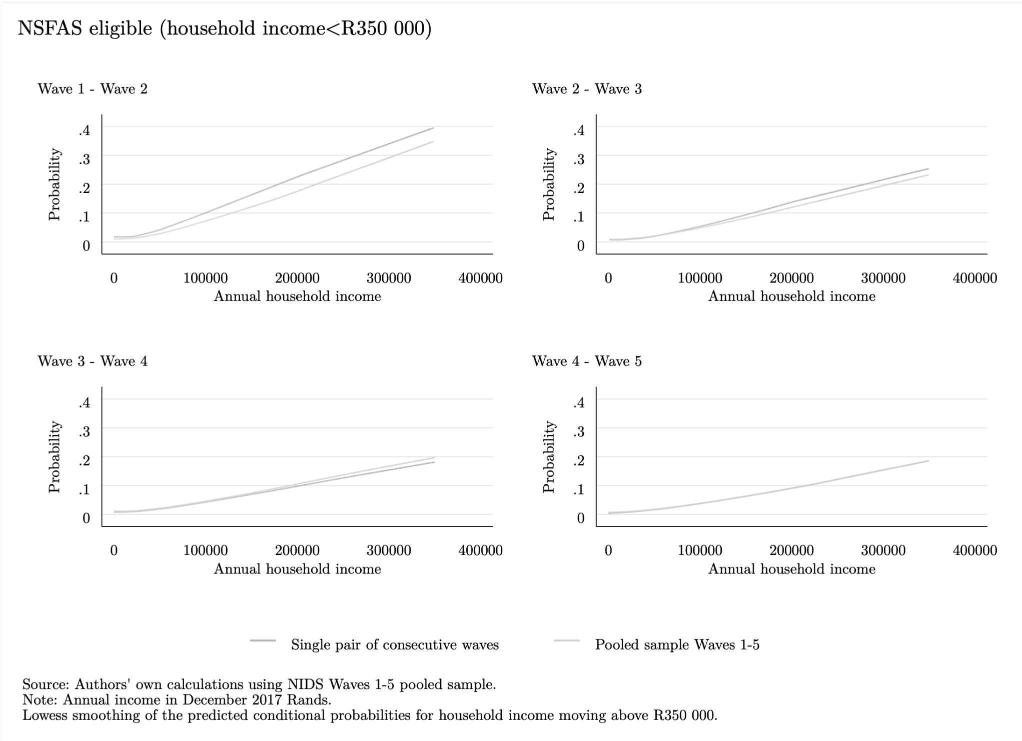
A further note on attrition is relevant here. Although the multivariate probit regression accounts for attrition directly, it is important to be aware that we include Temporary Sample Members [TSMs] in our base period sample. This ensures that we have a nationally representative sample, or as close to that as possible when using post-stratification weights in the base period. However, unless TSMs remain co-resident with a Continuing Sample Member [CSM] in the following wave, they are not actually tracked. Therefore, the inclusion of TSMs in the base period sample may misrepresent the type of attrition process from base period waves in which TSMs are present (i.e. Waves 2-4 – there are no TSMs in Wave 1), since TSMs will drop out of the study the moment that they cease to co-reside with a CSM. Although we include a binary variable in the retention equation for whether a respondent is a CSM or TSM, this only partly addresses this concern. The fact that we correct for both types of attrition in the pooled sample, but no TSMs drop out between Waves 1 and 2, may be why we observe better aligned distributions in the transitions between Waves 2-3, Waves 3-4, and Waves 4-5 compared to the jump in distributions between the Waves 1-2 and Waves 2-3 transitions.

Whether the size of certain groups is over- or under-estimated is thus difficult to disentangle, and concerns about attrition are not easily resolved, even with weights. As a robustness check for our funding classes defined from the pooled sample, we reclassify our strata based on the Wave 1-2 transition only, since Wave 1 is the most nationally representative base period, by design. In this case, the probability threshold between the vulnerable and stable missing middle is set at 37.20% – the observed rate of household income falling below the threshold in between Wave 1 and Wave 2.

Table 9 compares key households characteristics for the vulnerable and stable missing middle when these groups are estimated using the pooled sample (as in the main analysis) and the Wave 1-2 transition only. Only missing middle groups have been compared here, but a similar comparison can be made below the threshold. Table 15 in Appendix C shows the average of all characteristics of the household and household head for all funding classes as defined from the Wave 1-2 transition.

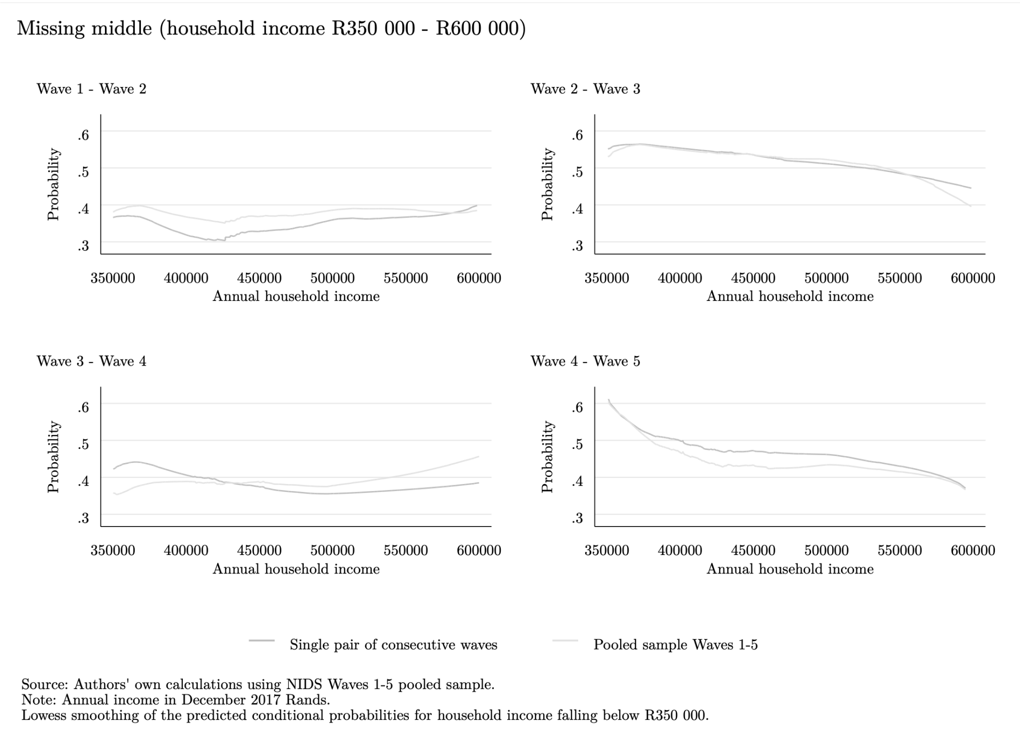
In general, the estimates of access to credit markets and financial assets are lower for both the stable and vulnerable missing middle when using the pooled sample rather than the Wave 1-2 transition only. This may owe to the fact that individuals who access these markets are more likely to leave the sample over time. Similarly, the estimated share of household income from the labour market is lower from the pooled sample for both groups. Despite these differences in estimated shares, the practical implications of the findings in the main analysis remain unchanged.

Figure 5: Distribution of predicted conditional probabilities of household income moving above the funding threshold



Source: Authors' own calculations using NIDS Waves 1-5 pooled sample.
Notes: Annual income in December 2017 Rands.
Lowess smoothing of the predicted conditional probabilities for household income moving above R350 000.

Figure 6: Distribution of predicted conditional probabilities of household income falling below the funding threshold



Source: Authors' own calculations using NIDS Waves 1-5 pooled sample.
Notes: Annual income in December 2017 Rands.
Lowess smoothing of the predicted conditional probabilities for household income moving above R350 000.

Table 9: Average characteristics of households and household heads in the missing middle (pooled transition vs. wave 1-2 transition)

	Vulnerable missing middle		Stable missing middle	
	Pooled transitions	Wave 1-2 transition	Pooled transitions	Wave 1-2 transition
Weighted share of respondents	2.17%	1.94%	2.41%	1.73%
(a) Characteristics of the head				
Race				
African	0.76	0.70	0.16	0.06
Coloured	0.13	0.12	0.13	0.10
Asian/Indian	0.05	0.12	0.06	0.04
White	0.06	0.06	0.65	0.79
Female household head	0.48	0.22	0.26	0.14
Household head is employed	0.64	0.74	0.79	0.76
Household head has a post-school qualification	0.37	0.43	0.71	0.54
(b) Household characteristics				
Per capita monthly income[a]	8 835.43	9 379.95	11 986.31	11 121.68
Share of household income from labour market	0.80	0.85	0.88	0.95
At least one resident has:				
Financial assets	0.94	0.99	0.94	0.99
Pension assets	0.31	0.27	0.52	0.51
Home loan/bond	0.24	0.29	0.53	0.66
Bank loan	0.34	0.38	0.28	0.26
Vehicle finance	0.30	0.47	0.35	0.34
Credit card	0.32	0.45	0.47	0.61
Household dependency ratio	0.63	0.63	0.49	0.51
Household employment rate[b]	0.57	0.55	0.74	0.66
Household is in an urban location	0.78	0.89	0.95	0.97

Source: Authors' own calculations using NIDS Waves 1 to 5 pooled sample.
Notes: Data are weighted using post-stratification weights from the base period.
[a] Value in December 2017 Rands. Labour market income reflects after tax income. [b] Of working age adults.

Appendix C Additional tables

Table 10: Multivariate probit: all equations (coefficient estimates)

	Transition eqn.		Initial	Retention
Base period covariate	Remain	Fall	conditions	eqn.
Characteristics of the household head				
Race (base: White)				
African	0.444**	0.702***	0.847***	0.557***
	(0.183)	(0.207)	(0.103)	(0.076)
Coloured	0.315*	0.451**	0.457***	0.446***
	(0.182)	(0.213)	(0.132)	(0.083)
Asian/Indian	0.548**	0.054	0.117	0.068
	(0.252)	(0.284)	(0.207)	(0.147)
Age of household head	-0.001	0.013**	-0.004	0.003**
	(0.003)	(0.006)	(0.003)	(0.001)
Household head is female (yes=1)	0.105*	0.245*	0.229***	0.011
	(0.061)	(0.135)	(0.060)	(0.027)
Household head is employed (yes=1)	0.123	0.218	0.008	0.048
	(0.088)	(0.155)	(0.077)	(0.037)
Highest education (base: Incomplete secondary)				
Missing	-0.537*	-2.488***	0.155	0.095
	(0.291)	(0.488)	(0.266)	(0.186)
No schooling	0.263*	-0.938**	0.504***	-0.128***
	(0.135)	(0.462)	(0.140)	(0.042)
Primary	0.052	-0.177	0.324***	-0.030
	(0.098)	(0.402)	(0.114)	(0.034)
Matric	-0.332***	-0.056	-0.184*	-0.183***
	(0.101)	(0.210)	(0.104)	(0.047)
Post-school qualification	-0.506***	-0.286	-0.434***	-0.185***
	(0.101)	(0.181)	(0.081)	(0.044)
Household characteristics				
Household employment rate	-0.379***	-0.367	-0.690***	-0.041
	(0.110)	(0.232)	(0.105)	(0.052)
At least one resident has financial assets (yes=1)	-0.038	-0.237	0.101	0.088***
	(0.092)	(0.266)	(0.087)	(0.028)
At least one resident has pension assets (yes=1)	-0.381***	-0.060	-0.405***	-0.109*
	(0.098)	(0.137)	(0.070)	(0.061)
Household member owns dwelling (yes=1)	-0.010	0.088	-0.539***	0.202***
	(0.079)	(0.207)	(0.088)	(0.034)
Household has livestock assets (yes=1)	0.288**	1.053***	0.095	0.096***
	(0.125)	(0.356)	(0.152)	(0.037)
Share of durables owned	-1.162***	1.358**	-2.559***	-0.010
	(0.377)	(0.551)	(0.266)	(0.128)
Number of rooms in house	-0.038***	-0.043*	-0.072***	0.002
	(0.014)	(0.024)	(0.013)	(0.007)
Household has access to electricity (yes=1)	-0.077	-0.274	0.175*	0.025
	(0.097)	(0.285)	(0.098)	(0.033)
Piped water on site (yes=1)	0.004	0.094	0.194*	0.068**
	(0.105)	(0.377)	(0.105)	(0.031)
Household has a flush toilet (yes=1)	-0.220*	-0.161	-0.341***	-0.033
	(0.133)	(0.243)	(0.115)	(0.036)
Number of household residents	-0.064**	0.024	-0.087***	-0.015
	(0.027)	(0.055)	(0.024)	(0.011)
Household dependency ratio	0.046	-0.061	0.170***	0.046**
	(0.057)	(0.147)	(0.061)	(0.019)
Number of children under 6	0.095	0.230*	0.074	0.044*

Continuation of Table 10				
	Transition eqn.		Initial	Retention
Base period covariate	Remain	Fall	conditions	eqn.
	(0.060)	(0.127)	(0.050)	(0.022)
Number of children aged 6-18	0.011	0.011	0.044	0.068***
	(0.036)	(0.093)	(0.043)	(0.017)
Location (base: Rural)				
Urban	-0.128	-0.026	-0.160	-0.101***
	(0.147)	(0.181)	(0.121)	(0.038)
Farm	0.035	-0.177	0.360	-0.093
	(0.175)	(0.370)	(0.255)	(0.061)
Year in time t (base: 2010)				
2012	0.169	0.054	-0.180**	0.444***
	(0.111)	(0.170)	(0.082)	(0.039)
2014	0.135	-0.338*	0.025	0.545***
	(0.090)	(0.187)	(0.083)	(0.037)
2017	0.252***	-0.229	0.078	0.470***
	(0.089)	(0.164)	(0.082)	(0.040)
Province (base: Western Cape)				
Eastern Cape	-0.245*	0.434	-0.168	-0.074
	(0.132)	(0.289)	(0.130)	(0.064)
Northern Cape	0.048	0.379	0.185	0.043
	(0.110)	(0.238)	(0.114)	(0.058)
Free State	-0.100	-0.320	-0.206	0.050
	(0.146)	(0.242)	(0.156)	(0.085)
KwaZulu-Natal	-0.119	0.119	-0.144	0.045
	(0.140)	(0.229)	(0.131)	(0.064)
North West	-0.213	0.871***	-0.365**	0.029
	(0.211)	(0.236)	(0.173)	(0.079)
Gauteng	-0.005	-0.032	-0.270**	0.019
	(0.129)	(0.174)	(0.110)	(0.063)
Mpumalanga	-0.232	-0.076	-0.504***	0.024
	(0.156)	(0.213)	(0.130)	(0.067)
Limpopo	-0.491***	-0.022	-0.508***	0.130*
	(0.156)	(0.251)	(0.149)	(0.069)
Constant	4.433***	-1.723**	3.733***	-1.346***
	(0.886)	(0.769)	(0.285)	(0.118)
Exclusion restrictions				
Mother's occupation (base: Missing)				
Agri/Elementary			0.066	
			(0.053)	
Professionals			-0.172***	
			(0.065)	
Semi-skilled/operator			0.014	
			(0.132)	
Clerks/sales workers			-0.044	
			(0.066)	
Craft/trade			0.108	
			(0.143)	
Never worked			-0.058	
			(0.052)	
Father's occupation (base: Missing)				
Agri/Elementary			0.169**	
			(0.072)	
Professionals			-0.172***	
			(0.059)	
Semi-skilled/operator			0.129	
			(0.086)	
Clerks/sales workers			0.155**	

Continuation of Table 10				
	Transition eqn.		Initial	Retention
Base period covariate	Remain	Fall	conditions	eqn.
			(0.068)	
Craft/trade			0.068	
			(0.060)	
Never worked			0.009	
			(0.054)	
Mother's education (base: Incomplete secondary)				
Missing			-0.159*	
			(0.082)	
No schooling			-0.063	
			(0.071)	
Primary			-0.077	
			(0.076)	
Matric			-0.030	
			(0.072)	
Post-school qualification			-0.083	
			(0.064)	
Father's education (base: Incomplete secondary)				
Missing			0.016	
			(0.072)	
No schooling			0.004	
			(0.075)	
Primary			-0.018	
			(0.070)	
Matric			-0.165**	
			(0.078)	
Post-school qualification			-0.275***	
			(0.078)	
Respondent characteristics				
CSM or a TSM? (CSM=1)				1.128***
				(0.027)
Very friendly during interview (yes=1)				0.134***
				(0.041)
Very attentive during interview (yes=1)				0.022
				(0.034)
Observations	116 462		116 462	116 462

Source: Authors' own calculations using NIDS Waves 1 to 5 pooled sample.
Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors are in parentheses. Standard errors are robust to arbitrary levels of intra-household correlation and the presence of repeated observations on the same individual. Data are weighted using post-stratification weights from the base period.

Table 11: Multivariate probit transition equation, fitted to each pair of waves independently (coefficient estimates)

Base period covariate	Wave 1	Wave 2	Wave 3	Wave 4
Characteristics of the household head				
Race (base: White)				
African	1.027***	1.183***	0.857***	0.839***
	(0.175)	(0.200)	(0.196)	(0.184)
Coloured	0.586***	0.881***	0.585***	0.573***
	(0.226)	(0.228)	(0.210)	(0.205)
Asian/Indian	0.391	-0.094	0.288	0.367
	(0.339)	(0.325)	(0.310)	(0.287)
Age of household head	0.003	-0.008	0.005	-0.000
	(0.005)	(0.005)	(0.005)	(0.004)
Household head is employed (yes=1)	0.200	-0.097	0.474***	0.023

Continuation of Table 11				
Base period covariate	Wave 1	Wave 2	Wave 3	Wave 4
	(0.137)	(0.155)	(0.149)	(0.111)
Household head is female (yes=1)	0.024	0.174	0.221**	0.279***
	(0.099)	(0.122)	(0.096)	(0.081)
Highest education (base: Incomplete secondary)				
Missing[a]	-0.928*	-	-0.967***	0.505
	(0.479)	-	(0.314)	(0.412)
No schooling	0.567*	0.447	-0.058	-0.158
	(0.314)	(0.298)	(0.205)	(0.206)
Primary	0.156	0.510***	-0.010	-0.316**
	(0.172)	(0.166)	(0.165)	(0.155)
Matric	-0.701***	-0.270	-0.481***	0.050
	(0.161)	(0.181)	(0.148)	(0.172)
Post-school qualification	-0.706***	-0.643***	-0.660***	-0.767***
	(0.166)	(0.151)	(0.153)	(0.119)
Household characteristics				
Household employment rate	-0.594***	-0.663***	-0.837***	-0.521***
	(0.179)	(0.217)	(0.180)	(0.153)
At least one resident has financial assets (yes=1)	0.024	-0.079	0.043	-0.133
	(0.130)	(0.148)	(0.138)	(0.138)
At least one resident has pension assets (yes=1)	-0.201	-0.541***	-0.505***	-0.523***
	(0.173)	(0.136)	(0.175)	(0.109)
Household member owns dwelling (yes=1)	0.020	-0.308*	-0.212	-0.360**
	(0.155)	(0.163)	(0.134)	(0.144)
Household has livestock assets (yes=1)	0.526*	0.290	0.090	0.661***
	(0.276)	(0.270)	(0.187)	(0.220)
Share of durables owned	-1.420***	-0.123	-2.565***	-3.438***
	(0.538)	(0.527)	(0.489)	(0.461)
Number of rooms in house	-0.060***	-0.082***	-0.061***	-0.050***
	(0.023)	(0.026)	(0.020)	(0.019)
Household has access to electricity (yes=1)	0.016	-0.107	0.061	0.162
	(0.148)	(0.183)	(0.198)	(0.183)
Piped water on site (yes=1)	0.073	-0.003	-0.004	0.216
	(0.205)	(0.219)	(0.209)	(0.160)
Household has a flush toilet (yes=1)	0.074	-0.705***	-0.432*	0.150
	(0.248)	(0.173)	(0.236)	(0.148)
Number of household residents	-0.179***	0.001	-0.020	0.025
	(0.058)	(0.053)	(0.038)	(0.034)
Household dependency ratio	-0.125	-0.072	0.343***	0.253***
	(0.086)	(0.098)	(0.114)	(0.074)
Number of children under 6	0.305***	0.166	-0.026	-0.166**
	(0.101)	(0.117)	(0.090)	(0.066)
Number of children aged 6-18	0.203***	-0.080	-0.088	-0.092
	(0.068)	(0.082)	(0.064)	(0.065)
Location (base: Rural)				
Urban	-0.061	0.062	-0.123	-0.336*
	(0.255)	(0.179)	(0.266)	(0.173)
Farm	0.533*	0.606*	0.261	0.247
	(0.320)	(0.321)	(0.404)	(0.277)
Province & time fixed effects	Yes	Yes	Yes	Yes
Pseudo log-likelihood	-27 761 214	-29 616 459	-29 246 099	-32 647 150
Model chi-squared (d.f.=129)	1 169.42	1 613.40	2 359.26	2 539.10
Number of clusters	5 908	8 275	10 662	13 472

Observations	22 584	27 219	31 288	35 371

Source: Authors' own calculations using NIDS Waves 1 to 5.
Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors are in parentheses. Standard errors are robust to arbitrary levels of intra-household correlation and the presence of repeated observations on the same individual. Data are weighted using post-stratification weights from the base period.
Column title reflects the base wave for each transition.
[a] Household head education missing is dropped in the Wave 2/Wave 3 transition because of collinearity (it predicts success perfectly). There are only 48 observations in the estimation sample of the Wave 2/Wave 3 transition that had household head education missing.

Table 12: Transitions between NSFAS classifications

Household income (time t-1)	Household income (time t)				
	Below R122 000	R122 000- R350 000	R350 000- R600 000	Above R600 000	Total
Below R122 000	86.27	12.64	0.87	0.22	100
R122 000 to R350 000	32.09	55.49	9.48	2.94	100
R350 000 to R600 000	14.51	32.06	29.63	23.80	100
Above R600 000	5.04	21.53	24.97	48.46	100
Total	69.38	22.55	4.67	3.40	100

Source: Authors' own calculations using NIDS Waves 1 to 5 pooled sample.
Notes: Data are weighted using post-stratified weights from the base period and corrected for panel attrition.

Table 13: Average characteristics of households and household heads by funding class (youth respondents)

	Eligible		Missing Middle			
	Persistent	Transient	Vulnerable	Stable	Elite	Total
Weighted share of youth	60.90%	32.00%	1.97%	2.12%	3.01%	100%
(a) Characteristics of the head						
Household head is employed	0.39	0.61	0.65	0.84	0.78	0.49
Age of household head	44.13	42.42	49.13	40.95	43.27	43.59
Female household head	0.63	0.47	0.46	0.26	0.31	0.56
Race						
African	0.93	0.76	0.79	0.21	0.3	0.84
Coloured	0.06	0.12	0.13	0.13	0.09	0.08
Asian/Indian	0.01	0.03	0.06	0.08	0.1	0.02
White	0.00	0.10	0.02	0.58	0.50	0.06
Highest education						
Missing	0.00	0.01	0.00	0.01	0.01	0.00
No schooling	0.18	0.02	0.03	0.01	0.00	0.12
Primary	0.27	0.13	0.11	0.00	0.03	0.21
Incomplete secondary	0.41	0.22	0.33	0.05	0.09	0.33
Matric	0.09	0.22	0.2	0.17	0.18	0.14
Post-school qualification	0.04	0.40	0.35	0.76	0.69	0.20
(b) Household characteristics						
Income and expenditure						
Per capita expenditure[a]	972.52	2 989.42	5 026.34	10 163.72	13 230.04	2 261.99
Per capita income[a]	1 336.04	3 428.71	8 748.65	12 609.23	22 682.6	3 034.02
Share of income from[b]:						
Labour market	0.43	0.71	0.81	0.91	0.84	0.55
Government grants	0.34	0.12	0.05	0.01	0.01	0.25
Investment income	0.01	0.02	0.03	0.03	0.05	0.02
Remittances	0.08	0.04	0.04	0.02	0.02	0.07
Subsistence agriculture	0.01	0.00	0.00	0.00	0.00	0.00
One-shot response (no source)	0.12	0.09	0.08	0.04	0.08	0.11
Assets, credit and infrastructure						
At least one resident has a:						
Home loan/bond	0.00	0.07	0.21	0.54	0.44	0.05
Bank loan	0.07	0.21	0.35	0.30	0.25	0.13
Study loan with a bank	0.00	0.02	0.05	0.04	0.04	0.01
Vehicle finance	0.00	0.08	0.28	0.40	0.41	0.05
Credit card	0.02	0.13	0.31	0.49	0.55	0.09
Store card	0.19	0.42	0.56	0.46	0.51	0.29
Financial assets	0.61	0.90	0.94	0.94	0.95	0.73
Pension assets	0.01	0.15	0.32	0.49	0.47	0.08
Household member owns dwelling	0.71	0.66	0.83	0.73	0.77	0.70
Household livestock assets	0.08	0.02	0.03	0.00	0.01	0.06
Number of rooms in house	3.64	4.80	5.58	5.84	6.54	4.18
Household has access to electricity	0.79	0.95	0.95	0.98	0.97	0.86
Piped water on site	0.66	0.92	0.93	0.99	0.96	0.76
Has a flush toilet	0.42	0.8	0.81	1.00	0.96	0.58
Household composition						
Number of household residents	5.19	5.25	6.04	3.70	4.22	5.17
Number of children under 6	0.84	0.63	0.84	0.35	0.34	0.74
Number of children aged 6–18	1.53	1.43	1.44	0.70	0.88	1.46
Number of elderly residents[c]	0.28	0.24	0.28	0.16	0.17	0.26
Household employment rate [d]	0.34	0.56	0.54	0.73	0.66	0.43
Location						
Traditional	0.43	0.20	0.21	0.04	0.06	0.34
Urban	0.51	0.77	0.76	0.94	0.93	0.62
Farm	0.06	0.02	0.03	0.02	0.02	0.05

Source: Authors' own calculations using NIDS Waves 1 to 5 pooled sample.
Notes: Data are weighted using post-stratification weights from the base period.
[a] Monthly value in Rands. [b] Excludes imputed rental income. Labour market income reflects after tax income. [c] Aged 60+. [d] Of working age adults.

Table 14: Average characteristics of households and household heads by threshold value and funding class (missing middle respondents)

	Threshold = 37.79%		Threshold = 50%		Threshold = 75%	
	Vulnerable	Stable	Vulnerable	Stable	Vulnerable	Stable
(a) Characteristics of the head						
Household head is employed	0.64	0.79	0.65	0.75	0.52	0.75
Age of household head	49.1	44.69	50.3	44.92	54.34	45.38
Female household head	0.48	0.26	0.51	0.29	0.59	0.32
Race						
African	0.76	0.16	0.83	0.24	0.94	0.35
Coloured	0.13	0.13	0.11	0.14	0.03	0.15
Asian/Indian	0.05	0.06	0.04	0.07	0.03	0.06
White	0.06	0.65	0.02	0.56	0.00	0.44
Highest education						
Missing	0.00	0.01	0.00	0.01	0.00	0.01
No schooling	0.03	0.01	0.03	0.01	0.05	0.01
Primary	0.09	0.00	0.13	0.00	0.26	0.01
Incomplete secondary	0.29	0.08	0.32	0.11	0.41	0.14
Matric	0.22	0.19	0.20	0.21	0.12	0.22
Post-school qualification	0.37	0.71	0.32	0.67	0.15	0.62
(b) Household characteristics						
Income and expenditure						
Per capita expenditure[a]	5 265.14	9 943.06	4 781.88	9 277.86	3 289.40	8 549.38
Per capita income[a]	8 835.43	11 986.31	8 218.88	11 691.68	6 233.05	11 283.60
Share of income from[b]:						
Labour market	0.80	0.88	0.79	0.88	0.69	0.88
Government grant	0.04	0.01	0.06	0.01	0.09	0.01
Other government income	0.01	0.00	0.02	0.00	0.03	0.00
Investment income	0.03	0.04	0.03	0.04	0.03	0.04
Remittances	0.03	0.01	0.03	0.01	0.05	0.01
Subsistence agriculture	0.00	0.00	0.00	0.00	0.01	0.00
One-shot response (no source)	0.08	0.05	0.08	0.06	0.10	0.06
Assets, credit and infrastructure						
At least one resident has a:						
Home loan/bond	0.24	0.53	0.18	0.51	0.09	0.46
Bank loan	0.34	0.28	0.32	0.31	0.29	0.32
Study loan with a bank	0.03	0.03	0.04	0.02	0.05	0.03
Vehicle finance	0.30	0.35	0.27	0.35	0.16	0.36
Credit card	0.32	0.47	0.31	0.45	0.23	0.43
Store card	0.55	0.44	0.54	0.46	0.48	0.49
Financial assets	0.94	0.94	0.93	0.94	0.92	0.94
Pension assets	0.31	0.52	0.29	0.48	0.24	0.45
Household member owns dwelling	0.86	0.78	0.86	0.79	0.87	0.8
Household owns livestock assets	0.04	0.00	0.05	0.00	0.09	0.00
Number of rooms in house	5.47	5.93	5.58	5.78	5.48	5.76
Access to electricity	0.94	0.97	0.92	0.97	0.91	0.96
Piped water on site	0.94	0.99	0.92	0.99	0.87	0.98
Has a flush toilet	0.82	1.00	0.77	0.99	0.67	0.96
Household composition						
Number of household residents	5.84	3.78	6.34	3.92	8.28	4.10
Household dependency ratio	0.63	0.49	0.64	0.51	0.79	0.51
Number of children under 6	0.83	0.33	0.98	0.35	1.48	0.40
Number of children aged 6–18	1.56	0.87	1.61	0.98	2.14	1.02
Number of elderly residents [c]	0.28	0.26	0.33	0.24	0.46	0.24
Household employment rate[d]	0.57	0.74	0.54	0.72	0.48	0.69
Location						
Traditional	0.19	0.04	0.23	0.05	0.31	0.08
Urban	0.78	0.95	0.75	0.93	0.66	0.90
Farm	0.03	0.01	0.02	0.02	0.03	0.02

Source: Authors' own calculations using NIDS Waves 1 to 5 pooled sample.
Notes: Data are weighted using post-stratification weights from the base period.
[a] Monthly value in Rands. [b] Share excludes imputed rental income. Labour market income reflects after tax income.
[c] Aged 60+. [d] Of working age adults.

Table 15: Average characteristics of households and household heads (Wave 1-2 transition)

	Eligible		Missing Middle			
	Persistent	Transient	Vulnerable	Stable	Elite	**Total**
Share of respondents	55.76%	36.31%	1.94%	1.73%	4.26%	100%
(a) Characteristics of the head						
Household head is employed	0.39	0.55	0.74	0.76	0.62	0.47
Age of household head	49.64	45.21	48.47	45.43	46.35	47.80
Female household head	0.53	0.40	0.22	0.14	0.11	0.45
Race						
African	0.96	0.67	0.70	0.06	0.20	0.80
Coloured	0.03	0.17	0.12	0.1	0.06	0.09
Asian/Indian	0.01	0.04	0.12	0.04	0.11	0.02
White	0.00	0.12	0.06	0.79	0.63	0.09
Highest education						
Missing	0.00	0.02	0.00	0.00	0.01	0.01
No schooling	0.30	0.04	0.00	0.00	0.01	0.18
Primary	0.39	0.17	0.05	0.00	0.02	0.28
Incomplete secondary	0.29	0.26	0.41	0.11	0.05	0.27
Matric	0.02	0.26	0.11	0.35	0.18	0.12
Post-school qualification	0.01	0.26	0.43	0.54	0.72	0.14
(b) Household characteristics						
Income and expenditure						
Per capita expenditure[a]	765.44	2 792.27	7 096.5	9 862.29	15 338.88	2 401.93
Per capita income[a]	888.87	2 765.53	9 379.95	11 121.68	25 392.68	2 954.99
Share of income from[b]:						
Labour market	0.39	0.74	0.85	0.95	0.87	0.56
Government grant	0.47	0.16	0.01	0.01	0.01	0.32
Other government income	0.01	0.01	0.07	0.00	0.00	0.01
Investment income	0.01	0.03	0.05	0.05	0.06	0.02
Remittances	0.06	0.05	0.02	0.00	0.05	0.05
Subsistence agriculture	0.01	0.00	0.00	0.00	0.00	0.01
One-shot response (no source)	0.03	0.01	0.00	0.00	0.00	0.02
Assets, credit and infrastructure						
At least one resident has a:						
Home loan/bond	0.01	0.12	0.29	0.66	0.50	0.09
Bank loan	0.04	0.16	0.38	0.26	0.15	0.10
Study loan with a bank	0.00	0.02	0.10	0.01	0.00	0.01
Vehicle finance	0.01	0.10	0.47	0.34	0.49	0.08
Credit card	0.02	0.16	0.45	0.61	0.67	0.12
Store card	0.12	0.36	0.68	0.51	0.44	0.24
Financial assets	0.4	0.75	0.99	0.99	0.96	0.57
Pension assets	0.01	0.11	0.27	0.51	0.39	0.08
Household member owns dwelling	0.82	0.70	0.83	0.76	0.88	0.78
Household owns livestock assets	0.16	0.03	0.00	0.00	0.02	0.10
Number of rooms in house	3.73	4.73	4.93	6.28	7.31	4.30
Access to electricity	0.69	0.9	0.97	0.93	0.98	0.79
Piped water on site	0.55	0.88	0.98	1.00	0.96	0.7
Has a flush toilet	0.33	0.74	0.95	0.99	0.95	0.53
Household composition						
Number of household residents	5.72	5.24	4.99	3.85	4.02	5.42
Household dependency ratio	0.96	0.66	0.63	0.51	0.57	0.82
Number of children under 6	1.02	0.68	0.45	0.41	0.40	0.85
Number of children aged 6–18	1.97	1.48	1.58	0.71	1.00	1.72
Number of elderly residents[c]	0.35	0.28	0.19	0.15	0.17	0.31
Household employment rate [d]	0.29	0.48	0.55	0.66	0.59	0.38
Location						
Traditional	0.50	0.20	0.09	0.02	0.05	0.36
Urban	0.42	0.77	0.89	0.97	0.94	0.59
Farm	0.08	0.03	0.02	0.01	0.01	0.05

Source: Authors' own calculations using NIDS Waves 1 to 2 pooled sample.
Notes: Data are weighted using post-stratification weights from the base period. [a] Monthly value in Rands. [b] Share excludes imputed rental income. Labour market income reflects after tax income. [c] Aged 60+. [d] Of working age adults.