How should we model a mortgage decision probability?

> Same mortgage application and zero-or-one outcome.

> Compare probability models and interpretable contrasts on the same observations.

Build a probability from an unrestricted index

First compute the index; then convert it to a bounded prediction.

> Index: $z = a + bX$

> Logistic examples: $z = -2, 0, 2 \Rightarrow p \approx 0.119, 0.500, 0.881$

Logit and probit keep fitted probabilities in range

Increasing S-shaped functions convert the index into a probability.

> Logit: $p(X) = \Lambda(z) = 1/(1 + e^{-z})$

> Probit: $p(X) = \Phi(z)$, area under a standard-normal curve left of z

> $z = a + bX$; both return probabilities in $(0, 1)$.

A logit coefficient changes log odds

Translate it before discussing probabilities.

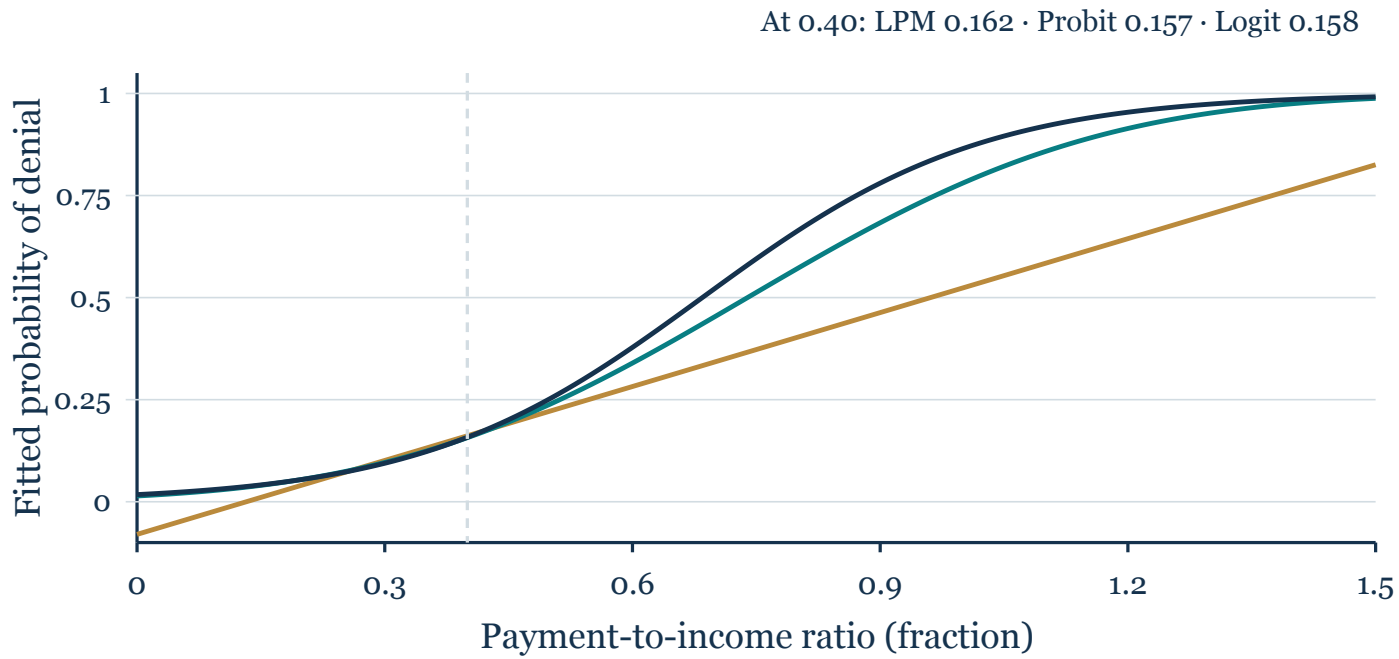
> Odds: $p/(1-p)$; log odds: $\log\{p/(1-p)\} = a + bX$

> $p = 0.20 \Rightarrow odds = 0.25$

> For $b = 0.5, \Delta X = 1$: $0.20 \to 0.292$, $0.50 \to 0.622$

Probit also uses an index coefficient

Its probability effect depends on the reference density.



Actual fitted LPM, probit and logit predictions from the same 2,380 applications. Payment-to-income is a fraction; these are descriptive fits.

- Probit: $p = \Phi(a + bX)$
- Local slope: $\partial p / \partial X = b\phi(a + bX)$
- Φ: cumulative area; ϕ: density height.

Compare two candidate models on observed events

Likelihood evaluates coefficient choices with the observed outcomes held fixed.

> Observed $(Y_1, Y_2) = (1, 0)$

> Candidate A: $0.8(1 - 0.2) = 0.64$

> Candidate B: $0.4(1 - 0.6) = 0.16$

Maximum likelihood chooses probabilities that support the observed events

Define the objective before interpreting the estimator.

> $L(\theta) = \prod_i p_i(\theta)^{Y_i}[1 - p_i(\theta)]^{1-Y_i}$

> $\ell(\theta) = \sum_i \{Y_i \log p_i + (1 - Y_i)\log(1 - p_i)\}$

> θ: coefficient list; $\prod$: multiply; $\sum$: add.

Compare the actual fitted probit and logit scales

Their coefficients are not directly comparable magnitudes.

```
. probit deny pir, vce(robust)

------------------------------------------------------------------------------
             |               Robust
        deny | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
         pir |   2.967907   .4653217     6.38   0.000     2.055893    3.879921
       _cons |  -2.194159   .1649761   -13.30   0.000    -2.517506   -1.870812
------------------------------------------------------------------------------
```

Predict probabilities at meaningful ratio values

State which covariates are changed or held fixed.

```
. margins, at(pir=(.3 .4 .5))

Adjusted predictions                                  Number of obs = 2,380
Model VCE: Robust

Expression: Pr(deny), predict()
1._at: pir = .3
2._at: pir = .4
3._at: pir = .5

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
         _at |
          1  |   .0961532   .0069017    13.93   0.000      .082626    .1096803
          2  |   .1569683   .0101166    15.52   0.000     .1371402    .1767965
          3  |   .2387884    .024485     9.75   0.000     .1907987    .2867781
------------------------------------------------------------------------------
```

A marginal effect is a local probability slope

For an indicator, use a discrete comparison instead.

> Logit: $\partial p/\partial X = b\,p(1-p)$

> Binary D: $p(D=1, X) - p(D=0, X)$

An average marginal effect averages individual slopes

The effect at the average covariates is another calculation.

```
. margins, dydx(pir)

Average marginal effects                                  Number of obs = 2,380
Model VCE: Robust

Expression: Pr(deny), predict()
dy/dx wrt:  pir

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
         pir |   .5664828   .0866687     6.54   0.000     .3966153    .7363504
------------------------------------------------------------------------------
```

$$AME = \frac{1}{n}\sum_i \left.\frac{\partial p(X)}{\partial X_j}\right|_{X=X_i}$$

Constructed logit: $X = (-2, 2) \Rightarrow AME \approx 0.105$

At mean $X = 0$, slope $= 0.25$.

A partial effect at average inputs is a different object

A nonlinear transformation does not commute with averaging.

> z_i: fitted index for applicant i; g: slope of the response function.

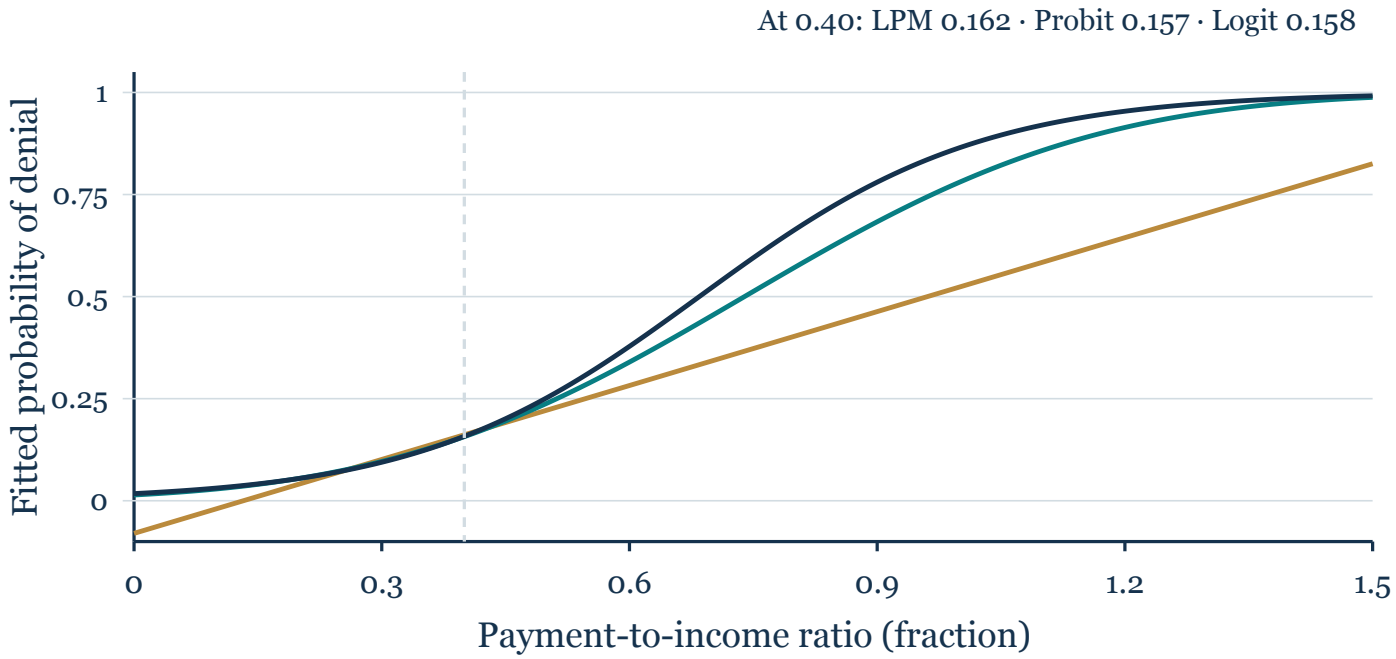
> $APE_j = AME_j = n^{-1} \sum_i g(z_i) b_j$.

> $PEA_j = g(z_{\text{mean}}) b_j$, where z_{mean} uses the average inputs.

> For an indicator, compare the fitted probabilities at zero and one.

A finite probability change can differ from a local approximation

Evaluate both endpoints when the proposed change is substantial.



Actual fitted LPM, probit and logit predictions from the same 2,380 applications. Payment-to-income is a fraction; these are descriptive fits.

Prediction performance needs more than a significance test

Check the probability model against its task.

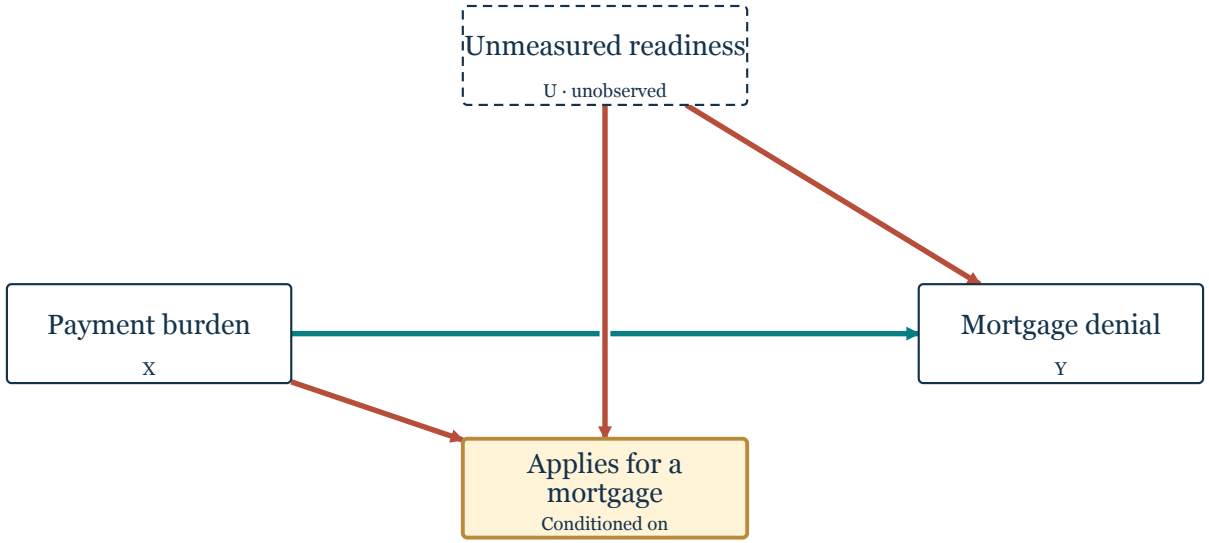
> Calibration: observed event frequencies agree with predicted probabilities.

> Classification: threshold probabilities to make event/non-event decisions.

> Assess new-data performance and the costs of different errors.

A fitted denial gap is not automatically a treatment effect

Selection and omitted applicant characteristics can remain.



1 open directed route · 1 open alternative · gold: conditioned on

Assumed causal structure · dashed boxes: unobserved · gold boxes: conditioned on

Write a compact applied research memo

Connect the question, data, model, assumptions and result.

> A useful memo identifies the decision or economic question, defines the observation and variables, explains the comparison and reports a meaningful magnitude with uncertainty.

> For the mortgage example, use a payment-burden contrast in percentage points and explain how it was obtained from the fitted model.

Compare probabilities on one common scale

Reproduce the evidence, then explain its economic meaning.

Practice

Add squared payment burden to the probit index. Recalculate the 0.30-to-0.40 fitted probability difference on the same observations and compare it with the original fit. Keep your commands, log, output and interpretation.

Compare probabilities on one common scale: explanation

A reproducible result needs both a calculation and an interpretation.

> Original linear-index probit contrast: 6.0815 percentage points.

> Quadratic-index probit contrast: 5.9786 percentage points.

> Same 2,380 applications; model flexibility does not establish a causal lending response.

Choose the method that addresses the empirical problem

A technical tool is useful when its role is explicit.

> Multiple regression accounts for stated observed differences.

> These roles can be combined, but none is a universal stamp of credibility.

Logit coefficient

Try the comparison before opening the explanation.

Practice

Why is a logit coefficient of 0.5 not a 50-percentage-point increase in probability?

Logit coefficient: explanation

Follow the calculation and interpret its units.

> It is a change in the log odds per unit of the regressor, conditional on the specification. The probability change depends on the starting index and other covariates. Report an explicit prediction contrast or marginal effect, meaning a local probability slope, to discuss probabilities.

A latent-variable representation fixes an otherwise arbitrary scale

Only the sign of the unobserved index is observed.

> $Y^\star = x\prime\beta + e, Y = 1\{Y^\star > 0\}.$

> Probit: $e \sim N(0,1)$; standard logit: logistic error variance $\pi^2/3$.

> Positive rescaling of the whole latent equation leaves the binary decision unchanged.

The Bernoulli log-likelihood makes the fitting criterion explicit

The observed outcome selects which probability enters each contribution.

> $p_i(\beta) = G(x_i'\beta), Y_i \in \{0, 1\}.$

> $L(\beta) = \prod_i p_i(\beta)^{Y_i}[1 - p_i(\beta)]^{1-Y_i}.$

> $\ell(\beta) = \sum_i \{Y_i \log p_i(\beta) + (1 - Y_i) \log[1 - p_i(\beta)]\}.$

Distinguish an average slope from a slope at average inputs

Try the comparison before opening the explanation.

Practice

For the mortgage probit model, explain what margins, dydx(pir) averages and what adding atmeans changes. Which reported quantity uses the observed covariate distribution? How should an indicator regressor be handled?

Distinguish an average slope from a slope at average inputs: explanation

Follow the calculation and interpret its units.

> The average marginal effect averages the model's input-specific local slopes over the observed applications. An atmeans calculation evaluates the slope at a single mean-input vector. Nonlinearity makes them generally different. Use a finite zero-to-one contrast for an indicator, with a clear reference population. Neither fitted probability comparison is automatically causal.

Average marginal effect

Try the comparison before opening the explanation.

Practice

Distinguish an average marginal effect from the marginal effect evaluated at the mean covariate vector.

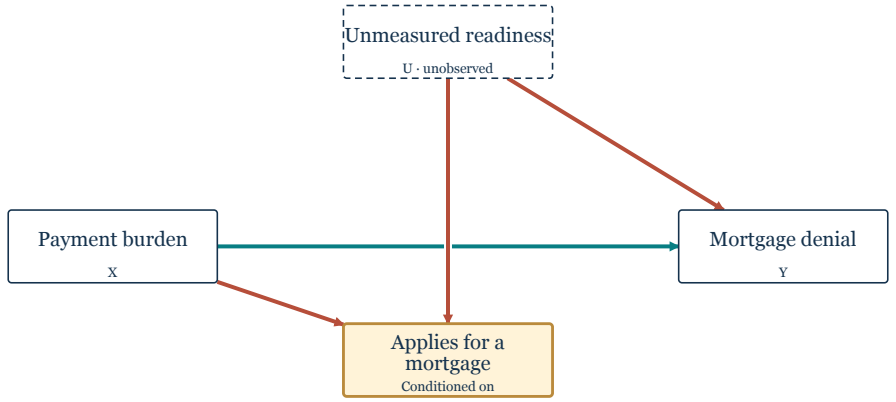
Average marginal effect: explanation

Follow the calculation and interpret its units.

> The AME computes an effect at each observation's covariates and averages those effects. The effect at means computes one effect at the vector of average covariates. In nonlinear models the two generally differ, and the average vector can describe no actual person.

DAG practice: selected applicants and prediction

Trace the routes before choosing a control.



1 open directed route · 1 open alternative · gold: conditioned on

Assumed causal structure · dashed boxes: unobserved · gold boxes: conditioned on

Practice

All observed records are mortgage applicants. In the diagram, payment burden and unmeasured readiness affect applying, while readiness also affects denial. Trace the route opened by selecting applicants. Does this make within-applicant prediction meaningless, or does it change the causal claim that needs defending?

DAG practice: selected applicants and prediction: explanation

Explain what conditioning changes in this stated graph.

> Selecting on application opens $X \rightarrow S \leftarrow U \rightarrow Y$, because application is a common consequence of burden and readiness. A causal comparison of burden can then mix its relationship with the outcome and selection through readiness. Predicting denial conditional on burden within the observed applicant population remains a legitimate descriptive task. A causal change in burden, or a conclusion about all potential borrowers, requires a clear target and additional identification and selection assumptions. A bounded logit or probit function does not itself close the path.