# Loan Denial Models and Marginal Effects

Prof. Conlon

2025-10-05

### Data Preparation

```
data("Hmda")

data <- Hmda %>%
  mutate(
    deny = 1 * (deny == "yes"),
    black = 1 * (black == "yes"),
    pbcr = 1 * (pbcr == "yes"),
    dmi = 1 * (dmi == "yes"),
    lvr_high = (lvr > 0.95),
    lvr_med  = (0.8 <= lvr) & (lvr <= 0.95)
  )

formula <- deny ~ black + dir + hir + lvr_med + lvr_high + ccs + mcs + pbcr + dmi
```

### Model Estimation

```
lpm    <- feols(formula, data = data)
```

```
## NOTE: 1 observation removed because of NA values (RHS: 1).
```

```
probit <- feglm(formula, data = data, family = binomial(link = "probit"))
```

```
## NOTE: 1 observation removed because of NA values (RHS: 1).
```

```
logit  <- feglm(formula, data = data, family = binomial(link = "logit"))
```

```
## NOTE: 1 observation removed because of NA values (RHS: 1).
```

### Regression Results

```
tab <- etable(
  lpm, probit, logit,
  title = "Loan Denial Models (LPM, Probit, Logit)",
  tex = TRUE,
  digits = 3,
  fitstat = ~pr2 + n,
```

```
  dict = c(
    "black" = "Black",
    "dir" = "Debt/Income",
    "hir" = "Housing/Income",
    "lvr_medTRUE" = "LTV: Medium",
    "lvr_highTRUE" = "LTV: High",
    "ccs" = "Consumer Credit",
    "mcs" = "Mortgage Credit",
    "pbcr" = "Public Bad Credit",
    "dmi" = "Denied Mortgage Ins"
  )
)

cat(tab)
```

```
writeLines(tab, "etable_output.tex")
```

### Average Marginal Effects for `black`

```
ame_lpm <- avg_slopes(lpm, variables = "black")
ame_probit <- avg_slopes(probit, variables = "black")
ame_logit  <- avg_slopes(logit,  variables = "black")

ame_tbl <- data.frame(
  Model = c("LPM AME (black)","Probit AME (black)", "Logit AME (black)"),
  Estimate = round(c(ame_lpm$estimate, ame_probit$estimate, ame_logit$estimate), 3),
  Std_Error = round(c(ame_lpm$std.error, ame_probit$std.error, ame_logit$std.error), 3)
)

kable(
  ame_tbl,
  format = "latex",
  booktabs = TRUE,
  caption = "Average Marginal Effects for `black`"
)
```

Now plot predicted probabilities

```
# Construct a grid of values for 'dir' (debt-to-income ratio)
dir_seq <- seq(min(data$dir, na.rm=TRUE),
               max(data$dir, na.rm=TRUE),
               length.out = 100)

# Create a dataset for prediction
# Set black = 0 and 1; other covariates fixed at their median
pred_data <- expand.grid(
  black = c(0, 1),
  dir = dir_seq,
```

Table 1: Loan Denial Models (LPM, Probit, Logit)

| Dependent Variable: | deny | | |
|---|---|---|---|
| Model: | (1) | (2) | (3) |
| | OLS | Probit | Logit |
| *Variables* | | | |
| Constant | -0.174*** | -2.96*** | -5.56*** |
| | (0.026) | (0.205) | (0.406) |
| Black | 0.081*** | 0.367*** | 0.657*** |
| | (0.017) | (0.097) | (0.177) |
| Debt/Income | 0.471*** | 2.58*** | 5.03*** |
| | (0.087) | (0.546) | (1.03) |
| Housing/Income | -0.069 | -0.328 | -0.405 |
| | (0.096) | (0.652) | (1.24) |
| LTV: Medium | 0.028** | 0.193** | 0.428*** |
| | (0.012) | (0.081) | (0.158) |
| LTV: High | 0.189*** | 0.779*** | 1.48*** |
| | (0.033) | (0.175) | (0.309) |
| Consumer Credit | 0.031*** | 0.153*** | 0.286*** |
| | (0.004) | (0.021) | (0.040) |
| Mortgage Credit | 0.019* | 0.134* | 0.258* |
| | (0.011) | (0.074) | (0.141) |
| Public Bad Credit | 0.200*** | 0.712*** | 1.25*** |
| | (0.023) | (0.118) | (0.205) |
| Denied Mortgage Ins | 0.701*** | 2.54*** | 4.53*** |
| | (0.041) | (0.284) | (0.554) |
| *Fit statistics* | | | |
| Pseudo $R^2$ | 0.51868 | 0.26407 | 0.26586 |
| Observations | 2,380 | 2,380 | 2,380 |

*IID standard-errors in parentheses*
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table 2: Average Marginal Effects for 'black'

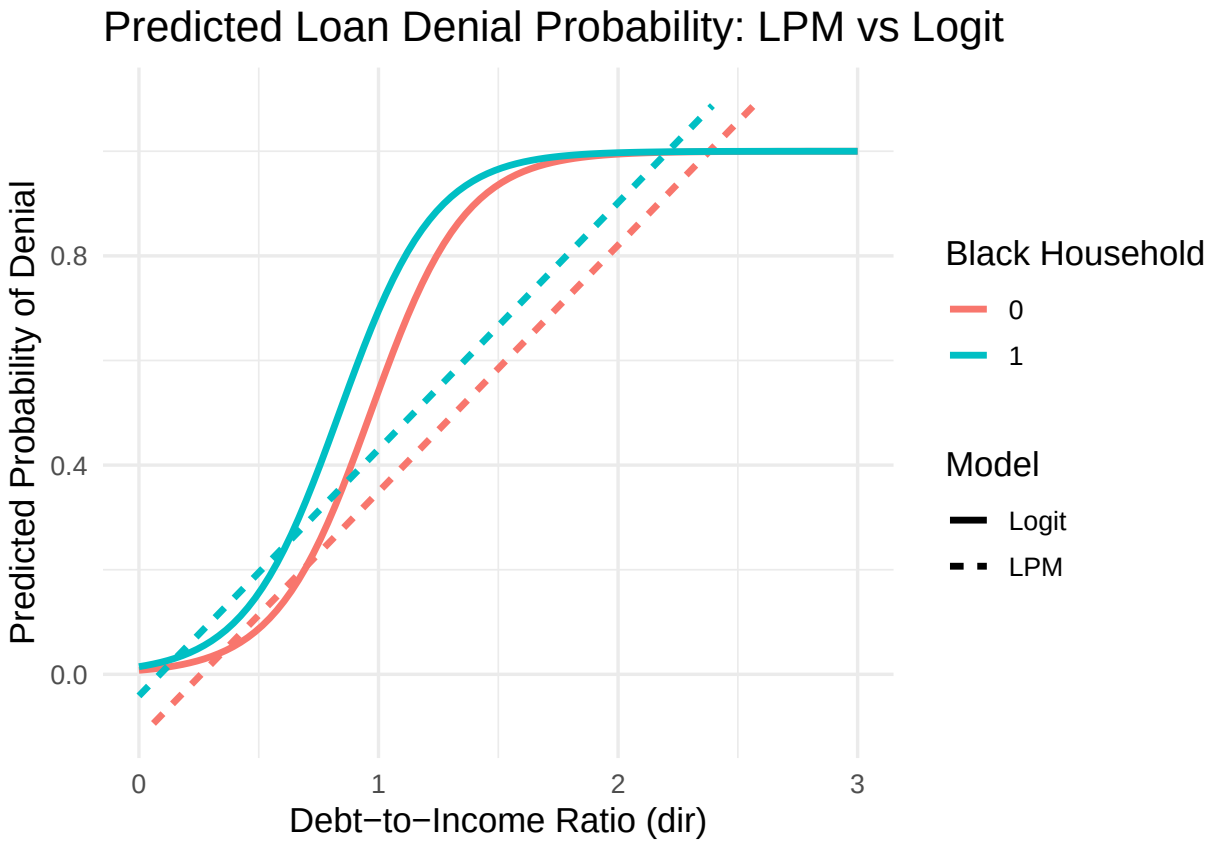
| Model | Estimate | Std_Error |
|---|---|---|
| LPM AME (black) | 0.081 | 0.017 |
| Probit AME (black) | 0.062 | 0.019 |
| Logit AME (black) | 0.057 | 0.018 |

```
  hir = median(data$hir, na.rm=TRUE),
  lvr_med = median(as.numeric(data$lvr_med), na.rm=TRUE),
  lvr_high = median(as.numeric(data$lvr_high), na.rm=TRUE),
  ccs = median(data$ccs, na.rm=TRUE),
  mcs = median(data$mcs, na.rm=TRUE),
  pbcr = median(as.numeric(data$pbcr), na.rm=TRUE),
  dmi = median(as.numeric(data$dmi), na.rm=TRUE)
)
# Get predicted probabilities
pred_lpm   <- predictions(lpm, newdata = pred_data) |> mutate(model = "LPM")
pred_logit <- predictions(logit, newdata = pred_data) |> mutate(model = "Logit")
pred_all <- bind_rows(pred_lpm, pred_logit)

# Plot
ggplot(pred_all, aes(x = dir, y = estimate, color = factor(black), linetype = model)) +
  geom_line(size = 1.2) +
  labs(
    x = "Debt-to-Income Ratio (dir)",
    y = "Predicted Probability of Denial",
    color = "Black Household",
    linetype = "Model",
    title = "Predicted Loan Denial Probability: LPM vs Logit"
  ) + ylim(-0.1, 1.1) +
  theme_minimal(base_size = 13)
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

```
## Warning: Removed 36 rows containing missing values or values outside the scale range
## (`geom_line()`).
```




```
ggsave('predicted_effects.pdf')
```

```
## Saving 6.5 x 4.5 in image
```

```
## Warning: Removed 36 rows containing missing values or values outside the scale range
## (`geom_line()`).
```