Big Data in Finance

Regularization, and Introduction to Mortgages

Tarun Ramadorai

Regularization

- One important tool that we didn't cover in great detail in the previous lectures is Regularization.

- Regularization refers to the process of introducing additional constraints into the optimization problem in order to get a more sensible solution, and in particular to avoid the problem of overfitting.

- This is one of the most important and useful methods in the machine learning toolkit, as we will see.

- Helps to solve a common problem: How can we make progress if there are many possible predictors of an outcome variable?

Linear Regression, Prediction, and Regularization

- Consider the standard OLS regression:

$$Y_i = \beta_0 + \sum_{k=1}^{K} X_{ik}\beta_k + \varepsilon_i = X_i^{'}\beta + \varepsilon_i.$$

- We usually estimate the parameters of this regression as:

$$\arg\min_{\beta_0,\beta_{1,..,}\beta_K} \sum_{i=1}^{N} \left(Y_i - \beta_0 - \sum_{k=1}^{K} X_{ik}\beta_k \right)^2 .$$

- Several regularization approaches can be used to modify this objective function. Why should we do so?

Linear Regression, Prediction, and Regularization

- Using the linear regression is useful as an intuitive explanatory device, but less good for prediction.

- For example, when $K \geq 3$, OLS is not admissible. (See Efron article in Scientific American in the reading list for an intuitive explanation of Stein's (1955) famous result).
  - In practice, this means that there are better predictors than OLS when $K \geq 3$.

- Another issue is what we might do when K is very large – for example, in the credit assignment, there are over 100 possible variables but this can be much higher if we consider nonlinear transformations of these variables.
  - In medical science, often millions.

Linear Regression, Prediction, and Regularization

- One possible problem when you have many RHS variables is that you increase the chance that they are highly correlated with one another – meaning that their magnitudes may be off because of collinearity.

- Another is that it can make testing restrictions or hypotheses very difficult especially if K is high relative to N, the sample size.

- Finally there's the obvious issue of interpretation being very hard when there are numerous variables on the RHS, making it difficult to figure out which are truly important.

Regularization - Best Subsets

- Clearly as we introduce additional regressors into OLS, we increase the goodness of fit – but there is the usual bias-variance tradeoff.

- There are several ways to introduce regularization constraints.

- One way is to limit the set of regressors explictly (say $k \subset K$). This is called "best subset selection"

- Computationally hard! The brute force way to do this is to simply fit all models which have k regressors. There will be $\binom{K}{k}$ of these models. Then select amongst them using R^2 or some other metric (cross-validated prediction error, perhaps).

Regularization - Stepwise Regressions

- A simpler approach is offered by stepwise regression. The procedure is:

1. For $k = 0, ..., K - 1$

    1.1 Consider all models with $k + 1$ regressors, i.e, those that augment the current model with one additional predictive variable.

    1.2 Pick the best of these models using your preferred criterion (say R^2).

    1.3 Either keep going until you hit a pre-specified number of regressors, or pick the model with the $k + 1$ that minimizes cross-validated prediction error, or adjusted R^2.

Hedge Fund Factors with Stepwise Regression

Agarwal and Naik (2004)

Table 4: Results with HFR Equally-Weighted Indexes

This table shows the results of the regression $R_t^i = c^i + \sum_{k=1}^{K} l_k^i F_{k,t} + u_t^i$ for the eight HFR indexes during the full sample period from January 1990 to June 2000 period. The table shows the intercept (C), statistically significant (at five percent level) slope coefficients on the various buy-and-hold and option-based risk factors and adjusted R^2 (Adj-R^2). The buy-and-hold risk factors are Russell 3000 index (RUS), lagged Russell 3000 index (LRUS)), MSCI excluding the US index (MXUS), MSCI Emerging Markets index (MEM), Fama-French Size and Book-to-Market factors (SMB & HML), Momentum factor (MOM), Salomon Brothers Government and Corporate Bond index (SBG), Salomon Brothers World Government Bond index (SBW), Lehman High Yield Composite index (LHY), Federal Reserve Bank Competitiveness-Weighted Dollar index (FRBI), Goldman Sachs Commodity index (GSCI) and the change in the default spread in basis points (DEFSPR). The option-based risk factors include the at-the-money and out-of-money call and put options on the S&P 500 Composite index ($SPC_{a/o}$ and $SPP_{a/o}$). For the two call and put option-based strategies, subscripts *a* and *o* refer to at-the-money and out-of-the-money respectively.

Event Arbitrage		Restructuring		Event Driven		Relative Value Arbitrage		Convertible Arbitrage		Equity Hedge		Equity Non-Hedge		Short Selling	
Factors	l	Factors	l	Factors	l	Factors	l	Factors	l	Factors	l	Factors	l	Factors	l
C	0.04	C	0.43	C	0.20	C	0.38	C	0.24	C	0.99	C	0.56	C	-0.07
SPP_o	-0.92	SPP_o	-0.63	SPP_o	-0.94	SPP_o	-0.64	SPP_a	-0.27	RUS	0.41	RUS	0.75	SPC_o	-1.38
SMB	0.15	SMB	0.24	SMB	0.31	MOM	-0.08	LRUS	0.10	SMB	0.33	SMB	0.58	RUS	-0.69
HML	0.08	HML	0.12	HML	0.12	SMB	0.17	SMB	0.05	HML	-0.08	MEM	0.05	SMB	-0.77
		LRUS	0.06	RUS	0.17	HML	0.08	MEM	0.03	GSCI	0.08			HML	0.40
		LHY	0.13	MEM	0.06	MXUS	0.04	SBG	0.16						
		FRBI	0.27												
		MEM	0.09												
Adj-R^2	44.04	Adj-R^2	65.57	Adj-R^2	73.38	Adj-R^2	52.17	Adj-R^2	40.51	Adj-R^2	72.53	Adj-R^2	91.63	Adj-R^2	82.02

Regularization - Best Subsets

- Can also approach this using backward stepwise regression. The procedure begins by estimating the full model with K regressors. Then:

1. For $k = K, K-1, ..., 1$

   1.1 Consider all models with $k-1$ regressors, i.e, those that delete one regressor at a time from the full model.

   1.2 Pick the best of these models using your preferred criterion (say R^2).

   1.3 Either keep going until you hit a pre-specified number of regressors, or pick the model with the k that minimizes cross-validated prediction error, or adjusted R^2.

Regularization - Best Subsets

- Note that both forward and backward stepwise regression are substantially less computationally intensive than best subset selection overall.
- Also worth noting that backward stepwise regression cannot be used if $K > N$, the sample size.
- Forward stepwise is the only solution in this case.
- Now let's move across to understanding two other important techniques for regularization that are commonly applied in machine learning. These are Ridge Regression and the LASSO.

LASSO in Returns Forecasting
Chinco, Clark-Joseph, Ye (2016)

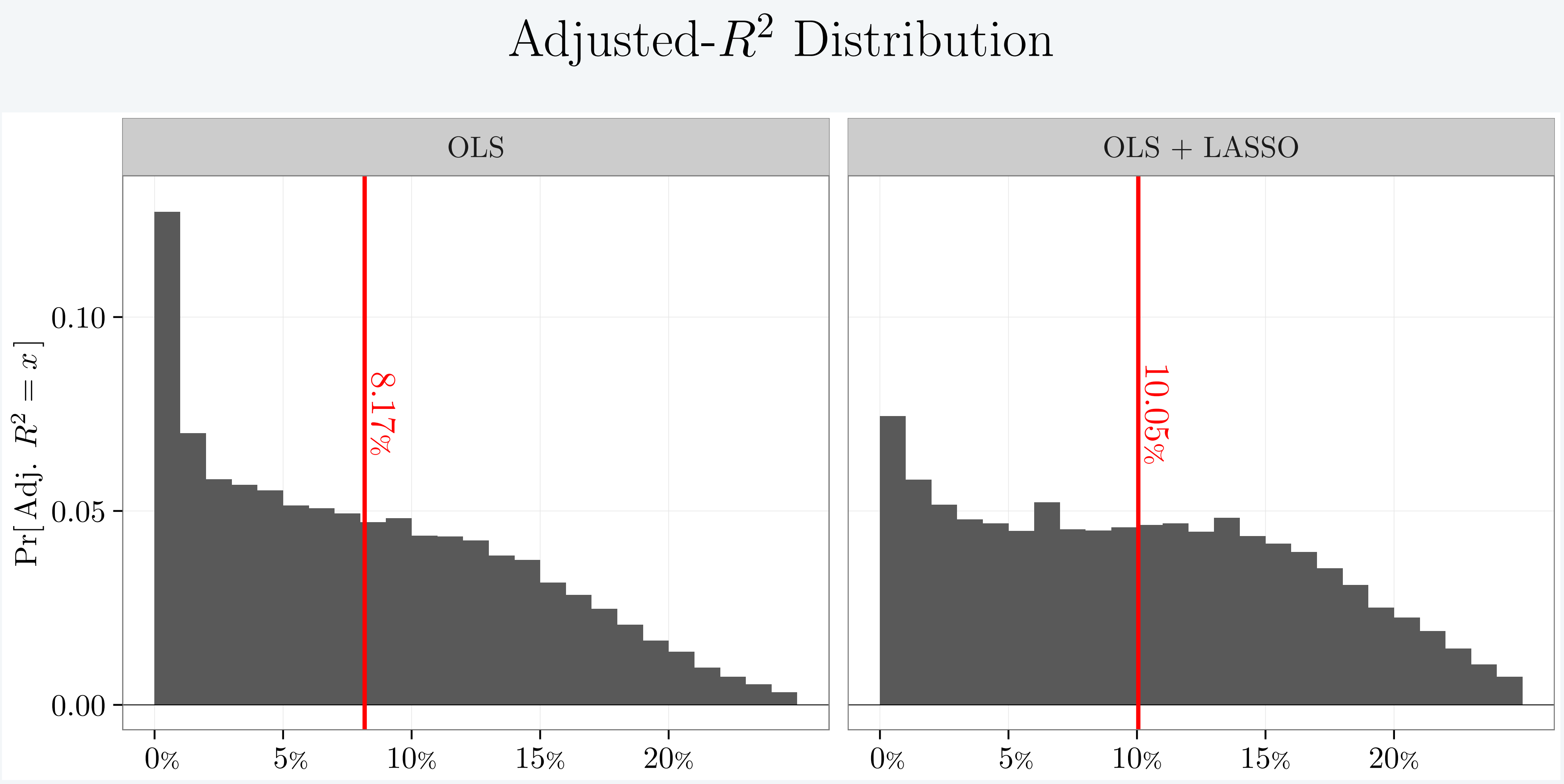


Figure 4: *Distribution of adjusted R^2s from the forecasting regressions in Equations* (4) *and* (8)*. Black bars: Probability that the adjusted R^2 from a single out-of-sample forecasting regression falls within a 1%-point interval. Red vertical line: Average adjusted R^2 from these regressions corresponding to the point estimates in the bottom row of Table 1. Left panel: Out-of-sample prediction made using OLS as in Equation* (4)*. Right panel: Out-of-sample predictions made using both OLS and the LASSO as in Equation* (8)*. Reads: "Including the LASSO's return forecast increases out-of-sample predictive power by* $^{10.05}/_{8.17} - 1 = 23\%$

Ridge Regression

- The Ridge Regression estimates parameters by minimizing the following objective function:

$$
\begin{aligned}
& \sum_{i=1}^{N}\left(Y_i-\beta_0-\sum_{k=1}^{K} X_{ik}\beta_k\right)^2+\lambda\sum_{k=1}^{K}\beta_k^2 \\
&= RSS+\lambda\sum_{k=1}^{K}\beta_k^2. \\
\hat{\beta}_{Ridge} &= (X'X+\lambda I_K)^{-1}(X'Y)
\end{aligned}
$$

Ridge Regression

- Here, $\lambda \geq 0$ is a tuning parameter to be estimated separately.
  - Can usually be selected using cross-validation in the training sample.

- Note what the objective function is saying: there is a penalty (l_2 norm) for a large number of large coefficients.

- This has the result of shrinking all the β_k estimates towards zero.

Ridge Regression

- In standard OLS, if we multiply one of the predictor variables by a constant c, the coefficient estimate simply scales by $\frac{1}{c}$. This means that for any scaled predictor, $X_k \hat{\beta}_k$ remains the same.

- However, because of the penalty function, this is not the case for the Ridge Regression. So it is best to first standardize the predictors before estimating (turn them into mean zero, variance one by dividing by in-sample standard deviation).

- Note also that the Ridge Regression will smoothly shrink the parameters to zero, i.e., if regressors are orthogonal to one another and normalized as above, all β_k coefficients will be shrunk by a factor of $\frac{1}{1+\lambda}$.

The LASSO

- As we saw earlier, est subset selection is problematic because it is computationally infeasible. The Ridge Regression, while computationally feasible, has the drawback that all predictors are generally selected (though shrunk).

- The LASSO does not have this drawback. The estimator is the solution to the optimization problem:

$$\sum_{i=1}^{N}\left(Y_i - \beta_0 - \sum_{k=1}^{K} X_{ik}\beta_k\right)^2 + \lambda \sum_{k=1}^{K} |\beta_k|$$
$$= RSS + \lambda \sum_{k=1}^{K} |\beta_k| .$$

- The estimator yields sparse models, which only involve a subset of variables.

Intuition for the LASSO
Chinco, Clark-Joseph, Ye (2016)

- Authors use LASSO to predict one-minute stock returns with the lagged returns of all the other stocks listed on the NYSE.

- Useful intuition for the LASSO, when the RHS variables in the LASSO regression are uncorrelated and have unit variance:
  - If $\hat{\beta}_k^{ols}$ is the OLS estimator, and $\hat{\beta}_k^{LASSO}$ the corresponding LASSO estimator, then:
$$\hat{\beta}_k^{LASSO} = \text{sign}(\hat{\beta}_k^{ols})[\max(0, \left|\hat{\beta}_k^{ols}\right| - \lambda)]$$
  - If OLS coefficient is estimated large (assume positive), then LASSO delivers similar estimate, since $\hat{\beta}_k^{LASSO} = \hat{\beta}_k^{ols} - \lambda \approx \hat{\beta}_k^{ols}$.
  - If the OLS coefficient is estimated small relative to λ, then $\hat{\beta}_k^{LASSO} = 0$.

Intuition for the LASSO

Chinco, Clark-Joseph, Ye (2016)

Relationship Between LASSO and OLS Estimates

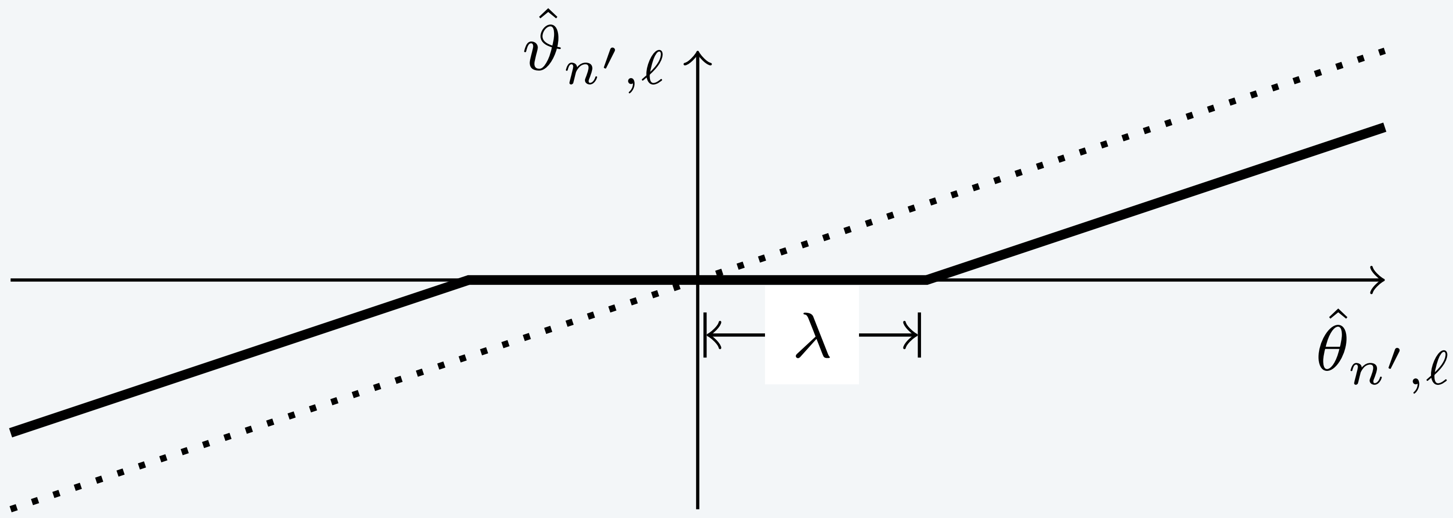
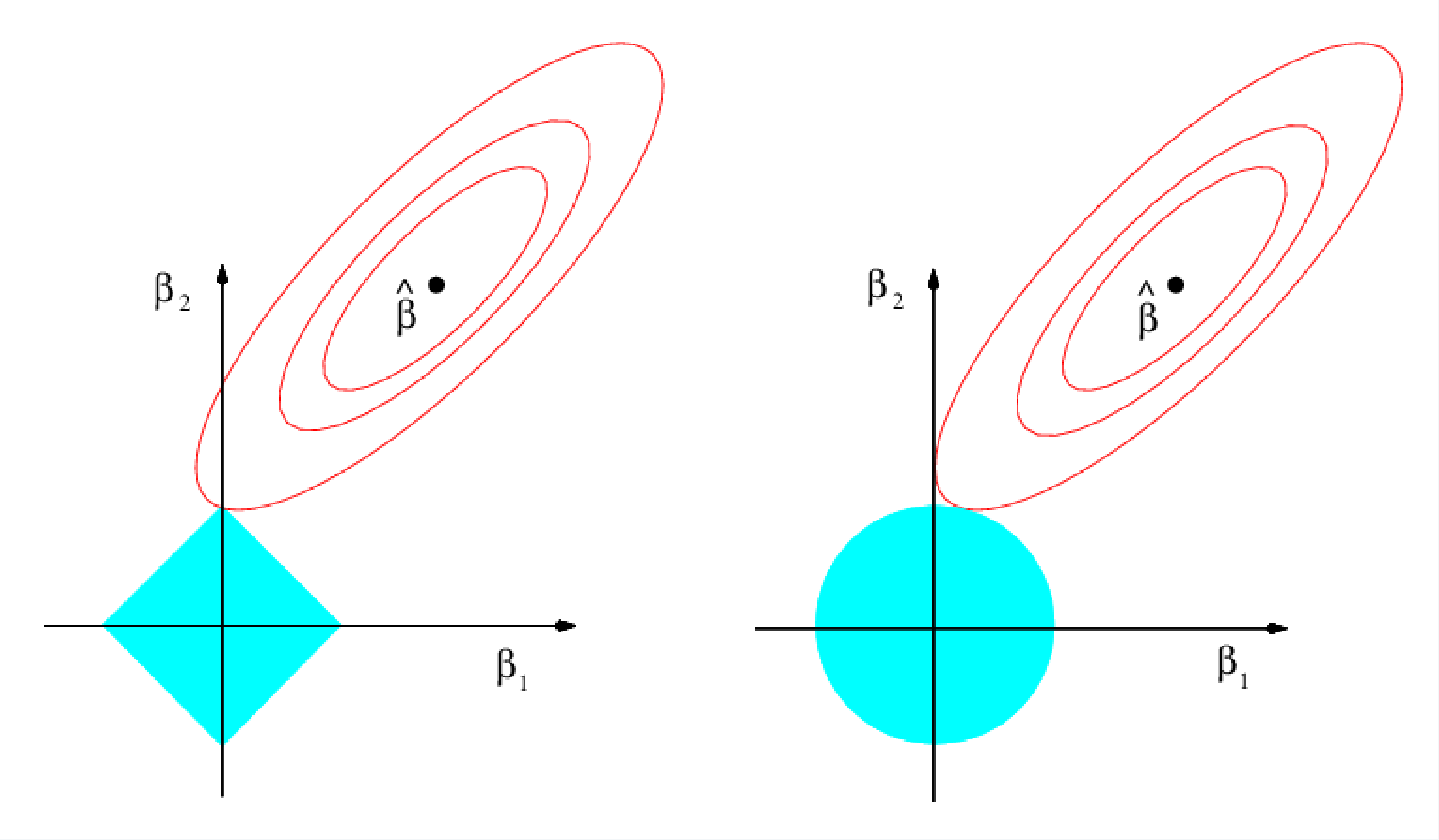


Figure 2: *x-axis: OLS-regression coefficient in an infinite sample. y-axis: Penalized-regression coefficient from the LASSO. Dotted: $x = y$ line. Reads: "If an OLS regression would have estimated a small coefficient value given enough data, $|\hat{\theta}_{n',\ell}| < \lambda$, then the LASSO will set $\hat{\vartheta}_{n',\ell} = 0$."*

Why Does LASSO Deliver Zero Coefficients?
Source: Intro to Stat. Learning



- Ellipses are iso-RSS curves; shaded are constraint regions for LASSO (left) and Ridge (right).

LASSO

- As you will see, the LASSO is an incredibly useful tool in a wide variety of contexts. You will use it in at least two cases in this course:
  - Credit scoring (including your assignment!).
  - Returns forecasting.

- But we will also consider how the LASSO helps with causal inference and not just prediction, later in the course.

- Now, a brief introduction to the topic of mortgages.

Why Mortgages?

- For the typical household, real estate is the single biggest asset, and mortgages are the single biggest liability.
- Relevant data for mortgages is certainly "big".
  - Enormous variety of mortgage contractual provisions and factors - choice is complex! (Guest speaker).
  - An important set of decisions through the mortgage lifecycle: Choice, Refinancing, Default.
  - Behavioural factors seem particularly important in this context given how emotive the choice of real estate is.
- Reasonably complex predictive empirical and theoretical models are needed to understand refinancing and mortgage default.
  - Well set up for using some of the techniques we have learned, and more that we will learn.

Mortgages: A Brief History

- Ancient Romans developed the legal concept of a *hypotheca*, a pledge of land to secure debt.
  - Germanic law called this pledge a "gage."
- In archaic French, turned into mort (dead) + gage (pledge).
- Why? Pledge is dead except in the instance of default.
- Terminology: Mortgagor is the borrower; Mortgagee is the lender.

Mortgages: Types and Variations

- Commercial (permanent versus bridge) and residential (US: conventional versus government insured).
- Repayment types:
  - Interest only versus principal-repayment mortgages.
  - Constant amortization versus constant payment mortgages.
  - Adjustable-rate mortgages versus fixed-rate mortgages (issue of pre-payment option).
- Other less common types:
  - Shared appreciation mortgage: lender receives share of price appreciation.
  - Reverse mortgage: borrower receives payments (often for older people needing income).
- Also relevant are foreclosure laws and other guarantees in case of default:
  - Priority of claims in foreclosure, costs of foreclosing.
  - Recourse versus non-recourse mortgages, Government guarantees.

Mortgage Choice

- Form of the contract has a number of different possible variations. Perhaps the most important features relate to how the interest, amortization, and payment are specified (next).
- Another important set of features has to do with the initial “fee” at origination is, and how this is traded off against the future interest payments on the mortgage.
- Yet another has to do with how prepayment penalties are specified on the mortgage contract - for fixed rate mortgages.
- Of course, the optimal choice for an individual depends on many factors, including (but not limited to):
  - expectations of variation in interest rates and realized movements in rates.
  - understanding of their own financial circumstances and evolution of these circumstances over time.
  - variety of contracts available, modulo financial constraints.

Mortgage Mechanics

- Rule 1: The interest owed in each payment equals the applicable interest rate times the outstanding principal balance (outstanding loan balance, or OLB) at the end of the previous period:

$$INT_t = (OLB_{t-1})r_t$$

- Rule 2: The principal amortized (paid down) in each payment equals the total payment (net of expenses and penalties) minus the interest owed:

$$AMORT_t = PMT_t - INT_t$$

- Rule 3: The outstanding principal balance after each payment equals the previous outstanding principal balance minus the principal paid down in the payment:

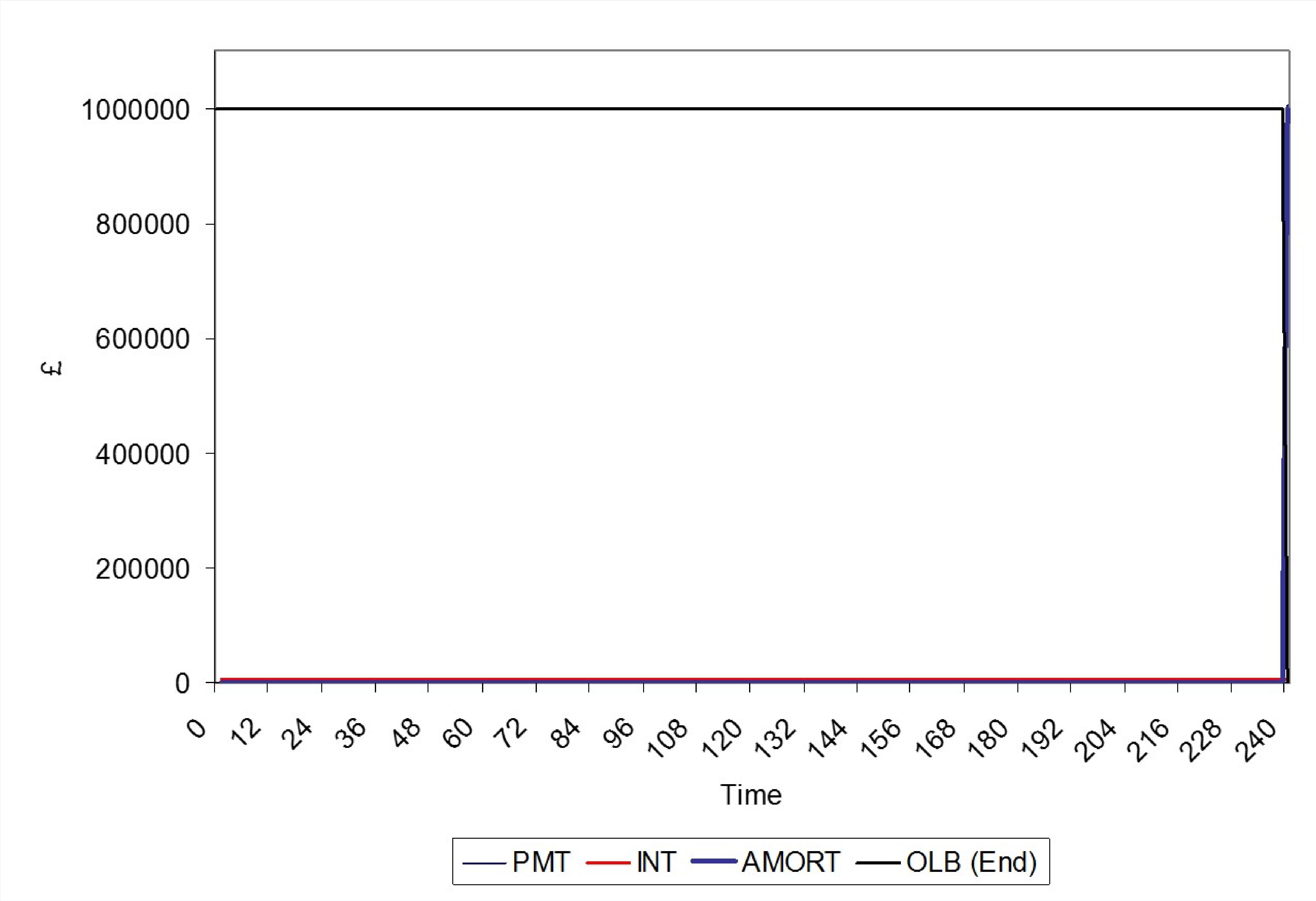
$$OLB_t = OLB_{t-1} - AMORT_t$$

- Rule 4: The initial outstanding principal balance equals the initial contract principal specified in the loan agreement:

$$OLB_0 = L$$

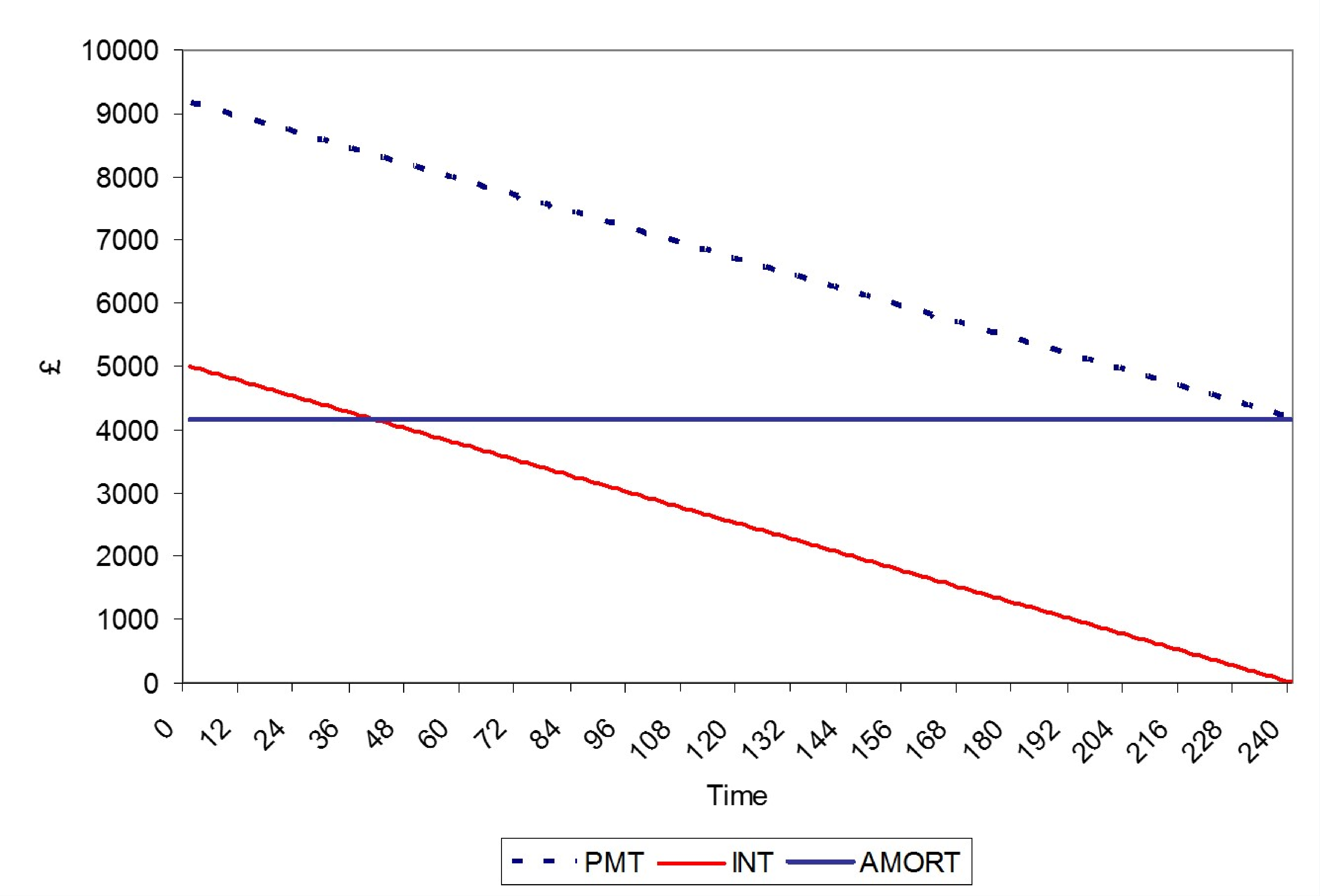
Interest-Only Loan

20 years, £1MM, 6% rate



Constant Payment Loan

20 years, £1MM, 6% rate



Pros and Cons

Interest-Only

Pros:

Minimizes mortgage payments.
Maximizes interest paid: potentially tax optimal.

Cons:

Higher interest paid overall.
Refinancing risk: "balloon payment" due at maturity.
Higher default risk because no amortization.

Constant Payment

Pros:

No balloon payment if fully amortizing.
If FRM, constant flat payments, easy to budget for.
Large initial interest payments: potentially tax beneficial.

Cons:

First-time buyers may not like constant large initial payments.
Constant nominal payments mean that buyer is subject to wealth risk because of inflation.

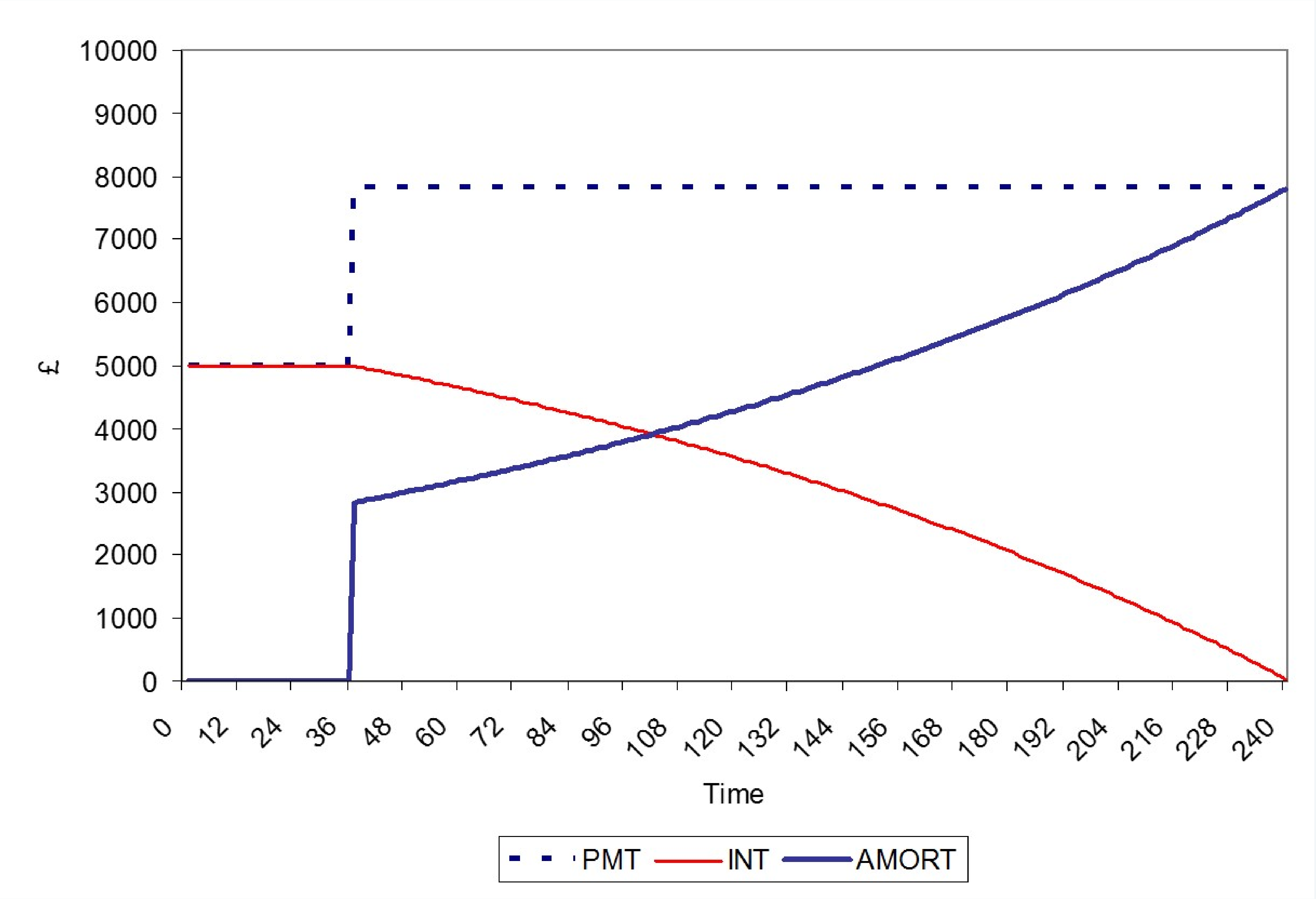
Choices, Choices!

- Can trade-off many features of the mortgage contract against one another.
  - Size of the monthly payment.
  - The speed of amortization.
  - Maturity.
  - Size of balloon payment: if the amortization period is longer than the mortgage maturity, this determines the size of the balloon payment.

Examples

20 year loan, interest-only for the first three years, with a reset at the end of this period. If you pay less interest in the first three years, then there is negative amortization over this period, and a countervailing ballooning of the payment stream.

Interest-Only with Reset After 3 Years



ARMs and FRMs

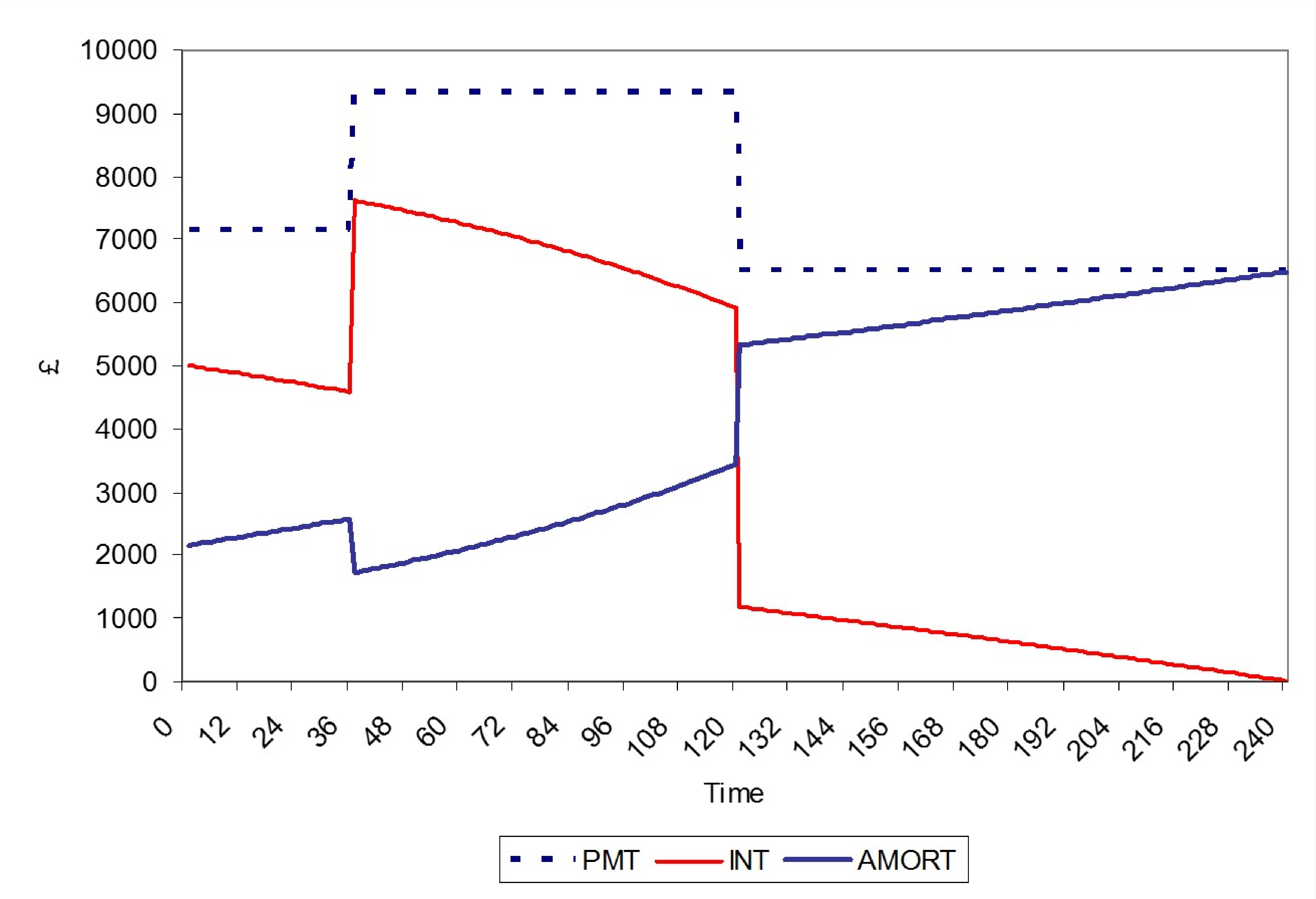
- The previous calculations assumed a constant interest rate (equal to 6%). This makes it a Fixed Rate Mortgage (FRM), with a constant interest rate over the life of the loan.

- In another type of contract, borrowers face interest rate risk. These are called Adjustable Rate Mortgages (ARMs).
  - Different countries have a different prevalence of mortgage contracts. For e.g., ARMs highly prevalent in the UK, FRMs in the US.

Examples

ARM with an interest rate of 6% initially that increases to 10% in year 4, and decreases to 2% in year 11

ARM: Interest Rate Risk



ARMs: Pros and Cons

Pros:

- Lower initial interest rate and payments – usually a “teaser” rate.
- Reduced interest rate risk for the lender.
- Some hedging for the borrower: Interest rates tend to rise during “good times”, and fall during “bad times”.

Cons:

- Non-constant payments, difficult to budget and administer.
- Interest rate risk for the borrower, meaning possibly greater default risk.
- Cons are mitigated by: adjustment intervals (longer intervals, less problems); interest rate (or payment) caps.

Rates and Yields

- Contract interest rate: interest rate that has been agreed between borrower and lender; fixed contractually.
- Yield = IRR on the loan. Changes over the life of the loan with changes in the market value of the loan.
  - Why might the two differ at origination?

- Loans have fees: 1% (one point) origination fee (= "prepaid interest" = "discount points" = "disbursement discount")
  - Example: £1 million loan, contract interest rate = 6%, 20 year maturity, 1% origination fee. What is the yield on the loan? Answer: 6.13% (annual).

$$0 = -990000 + \sum_{t=1}^{240} \frac{7164}{(1+Y)^t}.$$

Why Do Points and Fees Exist?

- Compensate brokers who find and filter applications for the lender.
- Pay back originators for overhead and administrative costs that occur up-front in the “origination” process.

- Bigger: To develop a “mortgage menu”, trading off up-front payment versus on-going monthly payment.
- To match borrower’s payment preferences.
- To screen borrowers: suppose that the borrower plans to terminate the contract early. Would she prefer higher up-front fees, or higher future mortgage payments?

Refinancing
More on this in the next lecture

- What would an individual that took out a FRM in at time 0 with an interest rate equal to 6% like to do if interest rates change at year 10 to 2%?
  - OLB = £645,314, PMT = £7,164.
  - If at the beginning of year 11 refinance and take out a 10-year FRM: PMT = £5,937.
- Who would lose out?
  - Solution: do not allow FRM borrowers to refinance (Realistic?).
  - Alternative solution: use pre-payment penalties.
- No incentive to pre-pay if pre-payment penalty = Present value of (£7,164 - £5,937) monthly over ten years = £133,301.
  - Is this realistic?
  - In practice long term FRMs have pre-payment penalties, but not so large. Need to figure out optimal exercise of the pre-payment option.
- Lenders should (and do) price in such an option in the initial contract interest rate/fees.

What's Next?

- You are probably beginning to get a sense of how complex the mortgage choice process is.

- There are also a bewildering array of supply-side attributes that need to be considered which limit lenders from offering particular deals (or make them keen to offer others).

- Many new firms have adopted entrepreneurial approach to solving the data problem that confronts households when choosing (and refinancing) mortgages.

- And next week, we will discuss models that harness the data on mortgages in various ways to predict default and refinancing.

Imperial College Business School

Imperial means
Intelligent Business