Haas School of Business **Professor Nancy Wallace**
Asset Securitization, MFE230M **Fall 2019**

Prepayment Modeling – Data Documentation

- For HW2 we will be estimating the coefficients of our prepayment model by using historical loan termination data. Note that these data are only used for estimating the parameters of the model. To calculate prepayment rates for the pools in the homework you will have to generate your own data using the Hull White interest rate model you calibrated in HW1.
- We have data representing non time varying covariates (static.csv) and a file representing time varying covariates (dynamic.csv). Both files represent the same set of loans.
- The data come from a sample of 24,183 Freddie Mac 30 year mortgages that were originated in 2003. Performance data are kept for the first 5 years of the mortgage.
- Both files contain the same variables
  - Column 1: Loan ID
  - Column 2: DTI (debt to income – at origination)
  - Column 3: LTV (loan to value – at origination)
  - Column 4: State
  - Column 5: UPB (unpaid balance – at origination)
  - Column 6: Ending UPB (unpaid balance at the end of the episode)
  - Column 7: Coupon gap on the loan defined as the contract rate on the mortgage minus the 3 month lag of the 10 year LIBOR rate
  - Column 8: Indicator variable equal to 1 if the episode is in May, June, July, or August
  - Column 9: Period marking the end of the episode
  - Column 10: Prepay Indicator (1 if loan became D120+ or prepaid during the episode)
  - Column 11: Period marking the start of the episode
  - ***Each row of the data file represents an episode.***
- With time non-varying covariates an episode is defined as the time from origination to termination, or to the end of the observation period if the loan does not terminate during the observation period. In this case we assume that covariates are only observed at the time of termination. The underlying assumption of this construction is that covariates do not change during the episode. So the coupon gap we observe refers to the difference between the loan coupon rate and the lagged 10-year LIBOR rate at termination. Similarly the indicator variable for summer is equal to 1 if termination occurs during May, June, July, or August.
- This is not a very good assumption as some episodes can span many years. We know that all the covariates will vary over the life of the loan, and more importantly the path of these covariates influence the termination decision.
- An alternative is to use time varying covariates. Here the duration of a loan is not treated as a single episode, but rather as a series of episodes. We assume that our data can be observed on a monthly basis and generate one episode per loan, per month of the duration period. A loan that was observed to terminate in period 15 with time non-varying covariates will now

have 14 episodes each with length of one month. Observations of the covariate are made at the beginning of each month. This is more intuitive than measuring at the end of the period.

- The variable of greatest importance with this construction is the variable indicating whether a loan has terminated (Column 4). This variable is equal to 1 only in the period of termination. All episodes prior to that have a value of 0.

Historical Interest Rate Data

- The cash flows of a mortgage backed security are interest rate dependent. In particular, to price bonds from a mortgage backed security we need to estimate the proportion of the pool that prepays in each period. Using the proportional hazard rate model we estimate coefficients that can be used to forecast the probability of prepayment in any given period.
- One of the key factors in the prepayment model is the coupon gap. To calculate this factor it is necessary to use an interest rate model such as Hull and White.
- We want to price the bonds as of the closing date - 08/30/2004 - so you will need to calibrate a Hull and White interest rate model and generate short rate paths that correspond to this date as you did for homework set one.