Machine Intelligence for Housing Finance

Kay Giesecke
Advanced Financial Technologies Laboratory
Stanford University
Housing

• Housing affects literally everyone

• Homeownership is a central pillar of the American Dream
  ~64% ownership rate
  Significant portion of household wealth tied up in homes

• US lenders process about 10 million mortgage applications every year

• The mortgage security market is massive ($10+ trillion)
Housing Market

1 Value of U.S. housing stock includes homes with and without underlying mortgages.
2 U.S. home equity is the difference between the value of the U.S. housing stock and the amount of U.S. single-family mortgage debt outstanding.

Mortgage Securities

Outstanding Public and Private Bond Market Debt – $40.5 Trillion

- Municipal Debt ($3.7T) 9%
- Treasury Debt 1 ($13.4T) 33%
- Agency Debt 2 ($1.9T) 5%
- Corporate Debt 6 ($8.4T) 21%
- Money Market Debt ($2.9T) 7%
- MBS 3 ($8.8T) 22%
- Asset-Backed Debt 4 ($1.4T) 3%

1 Interest-bearing marketable coupon public debt.
2 Includes Freddie Mac, Fannie Mae, Federal Home Loan Banks, Farmer Mac, the Farm Credit System, and federal budget agencies (e.g. TVA).
3 Includes Ginnie Mae, Fannie Mae and Freddie Mac mortgage-backed securities and CMOs, CMBS and private-label MBS/CMOs.
4 Includes auto, credit card, home equity, manufacturing, student loans and other USD-denominated CDOs are also included.
5 Includes commercial paper, bankers acceptances and large time deposits.
6 Includes all non-convertible debt, MTNs and Yankee bonds, but excludes CDs and federal agency debt.

Note: Percentages may not add up to 100% due to rounding.
Technology

• Yet housing finance technology is still stuck in the 20th century
  Paper based (~2000 pages per app), manual, slow (~45 days), costly
  Grid-based underwriting: bad decisions (miss good risks, approve bad ones)

• Some fintech startups (Blend, Roostify, …) are building 21st century technologies that improve origination efficiency and customer experience
Machine Intelligence

• Replacing the grid:
  Intelligent automation (underwriting, pricing, insurance, servicing, etc.)
  Transparent and efficient housing security markets

• Could generate potentially significant societal benefits
  Expand access to mortgage credit
  Reduce the risk of financial crises
  Improve capital allocation

• How? *Leverage data to understand borrower behavior*
“Man is the Measure of All Things.”

–Protagoras of Abdera, circa 490 BC
Borrower Behavior

Post-Origination
- Loan- and pool level risk
- Risk capital
- Policy (GSEs, etc.)
- Systemic risk

Servicization
- Structuring
- Rating
- Trading

Servicing
- Optimal actions
- Retention
- Servicing rights valuation

Application
- Who is in the market
- Prioritization (app to funding)
- Hedging (rate lock to close)

Underwriting
- KYC, Fraud
- Pre-approval
- Risk scoring
- Pricing
- Collateral valuation

Insurance
- Pricing
Role of GSEs

- U.S. Residential Mortgage Market
- Mortgage Securitization
- Mortgage-backed Securities
- Global Capital Markets
- Mortgage Investments
- Debt Securities
Pass-through MBS

Pass-through single family MRBs

- Bondholders (2.5% coupon)
  - Principal repayment, prepayment, and 2.5% interest
  - Bond Proceeds

- Trustee
  - MBS Interest Residual (0.5%)
  - Guarantees full and timely MBS payment

- HFA
- GSEs, GNMA

- Mortgage Servicer/Lender (Servicing Fee 0.25%)
  - Mortgage Loan (3.5%)
  - Monthly Payments
  - Guarantee Fee (0.25%)
  - Monthly Payments

- Homeowner
  - Mortgage Loan (3.5%)
  - Monthly Payments
Tranching

Sample Subprime RMBS Payments

<table>
<thead>
<tr>
<th>Monthly Mortgage Payments</th>
<th>REMIC Trust</th>
<th>Accounts</th>
<th>Interest Payments</th>
<th>Principal Payments</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1 M2 M3 M4 M5 M6 M7 M8 M9 M10</td>
<td>$</td>
<td>$ I</td>
<td>'AAA' L+% or NetWAC</td>
<td>'AAA' Scheduled Principal &amp; Prepayments</td>
</tr>
<tr>
<td>M11 M12 M13 M14 M15 M16 M17 M18 M19 M20</td>
<td></td>
<td></td>
<td>'AA' L+% or NetWAC</td>
<td>'AA'</td>
</tr>
<tr>
<td>M21 M22 M23 M24 M25 M26 M27 M28 M29 M30</td>
<td></td>
<td></td>
<td>'A' L+% or NetWAC</td>
<td>'A'</td>
</tr>
<tr>
<td>M31 M32 M33 M34 M35 M36 M37 M38 M39 M40</td>
<td></td>
<td></td>
<td>'BBB' L+% or NetWAC</td>
<td>'BBB'</td>
</tr>
<tr>
<td>M41 M42 M43 M44 M45 M46 M47 M48 M49 M50</td>
<td></td>
<td></td>
<td>'BBB' L+% or NetWAC</td>
<td>'BBB'</td>
</tr>
<tr>
<td>M51 M52 M53 M54 M55 M56 M57 M58 M59 M60</td>
<td></td>
<td></td>
<td>Residual Excess Interest</td>
<td>Residual</td>
</tr>
<tr>
<td>M61 M62 M63 M64 M65 M66 M67 M68 M69 M70</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M71 M72 M73 M74 M75 M76 M77 M78 M79 ... M 2000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M1 M2 M3 M4 M5 M6 M7 M8 M9 M10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M11 M12 M13 M14 M15 M16 M17 M18 M19 M20</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M21 M22 M23 M24 M25 M26 M27 M28 M29 M30</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M31 M32 M33 M34 M35 M36 M37 M38 M39 M40</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

'P' for Principal payments, 'I' for Interest payments.
CDO-Squared

Risk Profile of Subprime Mortgage Loans

- Low Borrower Down Payment
- Good Borrower Credit
  - AAA: 81%
  - AA: 11%
  - A: 4%
- Bad Borrower Credit
  - BBB: 3%
  - BB, NR: 1%, not in all deals

Subprime Mortgage Bonds

High Grade ABS CDO

- Senior AAA: 88%
- Junior AAA: 5%
- AA: 3%
- A: 2%
- BBB: 1%
- NR: 1%

Mezz ABS CDO

- Senior AAA: 62%
- Junior AAA: 14%
- AA: 8%
- A: 6%
- BBB: 6%
- NR: 4%

CDO²

- Senior AAA: 60%
- Junior AAA: 27%
- AA: 4%
- A: 3%
- BBB: 3%
- NR: 2%

Other credit support: Excess Spread, Over-collateralization
MSB Issuance

MBS Issuance Volume
$ in Trillions

<table>
<thead>
<tr>
<th>Year</th>
<th>Freddie Mac</th>
<th>Fannie Mae</th>
<th>Ginnie Mae</th>
<th>Private Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006</td>
<td>$2.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2007</td>
<td>$1.9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2008</td>
<td>$1.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2009</td>
<td>$1.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2010</td>
<td>$1.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td>$1.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2012</td>
<td>$1.7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2013</td>
<td>$1.6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2014</td>
<td>$0.9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2015</td>
<td>$1.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>YTD 2016</td>
<td>$0.6</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Enterprises & Ginnie Mae
- 2006: 44%
- 2007: 62%
- 2008: 95%
- 2009: 97%
- 2010: 96%
- 2011: 98%
- 2012: 99%
- 2013: 98%
- 2014: 96%
- 2015: 96%
- YTD 2016: 98%

Private Label
- 2006: 56%
- 2007: 38%
- 2008: 5%
- 2009: 3%
- 2010: 4%
- 2011: 2%
- 2012: 1%
- 2013: 2%
- 2014: 4%
- 2015: 4%
- YTD 2016: 2%

1 2016 data as of June 30, 2016.
Source: Inside Mortgage Finance.
Unprecedented Data

  - CoreLogic, Fannie Mae, Freddie Mac, Wells Fargo, etc.
- Property-level sales transactions 2000-2014 (~100 million sales)
  - 500 million county registrar of deed records
- Property-level tax records 2000-2014 (~150 million parcels)
  - County tax assessor
- Zip-code level economic, financial and demographic data
  - BLS, BEA, IRS, Fed, Zillow, Powerlytics, etc.
- Borrower-level bank account transaction data (Yodlee, Intuit)
- Borrower-level credit history data (bureaus)
- Borrower-level social media activity
- Property-level images (Zillow, StreetView, drone, satellite)
CoreLogic Data

- 120 million prime and subprime mortgages originated across the US between 1995 and 2014
  - Extensive loan- and borrower-level features at origination
  - Monthly performance update
- Data for local, regional, national economic factors from BLS, BEA, Zillow, FHFA, etc.
- 3.5 billion monthly observations, each characterized by ~300 features
# Prime vs. Subprime Loans

<table>
<thead>
<tr>
<th>Feature</th>
<th>Mean</th>
<th>Median</th>
<th>25%</th>
<th>75%</th>
</tr>
</thead>
<tbody>
<tr>
<td>FICO</td>
<td>720</td>
<td>730</td>
<td>679</td>
<td>772</td>
</tr>
<tr>
<td>LTV</td>
<td>74</td>
<td>79</td>
<td>63</td>
<td>90</td>
</tr>
<tr>
<td>Interest rate</td>
<td>5.8</td>
<td>5.8</td>
<td>4.9</td>
<td>6.6</td>
</tr>
<tr>
<td>Balance</td>
<td>190,614</td>
<td>151,353</td>
<td>98,679</td>
<td>238,000</td>
</tr>
</tbody>
</table>

Table 1: Prime mortgages

<table>
<thead>
<tr>
<th>Feature</th>
<th>Mean</th>
<th>Median</th>
<th>25%</th>
<th>75%</th>
</tr>
</thead>
<tbody>
<tr>
<td>FICO</td>
<td>634</td>
<td>630</td>
<td>580</td>
<td>680</td>
</tr>
<tr>
<td>LTV</td>
<td>74</td>
<td>80</td>
<td>68</td>
<td>90</td>
</tr>
<tr>
<td>Interest rate</td>
<td>7.8</td>
<td>7.8</td>
<td>6.3</td>
<td>9.6</td>
</tr>
<tr>
<td>Balance</td>
<td>160,197</td>
<td>124,000</td>
<td>68,850</td>
<td>210,000</td>
</tr>
</tbody>
</table>

Table 2: Subprime mortgages
Foreclosures by County (2006)

Cumulative Foreclosures Per 1000 Population

- [0,1]
- (1,5]
Foreclosures by County (2012)

Cumulative Foreclosures Per 1000 Population:

- [0,1]
- (1,5]
- (5,10]
- (10,20]
- (20,30]
- (30,50]
- (50,80]
<table>
<thead>
<tr>
<th></th>
<th>Current</th>
<th>30</th>
<th>60</th>
<th>90+</th>
<th>Foreclosure</th>
<th>REO</th>
<th>Paid Off</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current</td>
<td>97</td>
<td>1.4</td>
<td>0</td>
<td>0</td>
<td>.001</td>
<td>0</td>
<td>1.6</td>
</tr>
<tr>
<td>30 days</td>
<td>34.6</td>
<td>44.6</td>
<td>19</td>
<td>0</td>
<td>.004</td>
<td>.003</td>
<td>1.8</td>
</tr>
<tr>
<td>60 days</td>
<td>12</td>
<td>16.8</td>
<td>34.5</td>
<td>34</td>
<td>1.6</td>
<td>.009</td>
<td>1.1</td>
</tr>
<tr>
<td>90+ days</td>
<td>4.1</td>
<td>1.4</td>
<td>2.6</td>
<td>80.2</td>
<td>10</td>
<td>.3</td>
<td>1.3</td>
</tr>
<tr>
<td>Foreclosure</td>
<td>1.9</td>
<td>.3</td>
<td>.1</td>
<td>6.8</td>
<td>87</td>
<td>2.5</td>
<td>1.3</td>
</tr>
<tr>
<td>REO</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>Paid off</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100</td>
</tr>
</tbody>
</table>
Nonlinear Effects
Nonlinear Effects

The graph illustrates the prepayment rate as a function of the maximum difference between the initial interest rate and the national mortgage rate, with a floor of 0. The graph compares Prime Mortgages (blue line) and Subprime Mortgages (red line).
The conditional probability that the $n$-th loan transitions from its state $U^n_t$ at time $t$ to state $u$ at time $t+1$ is

$$P(U^n_{t+1} = u \mid F_t) = h_\theta(u, X^n_t)$$

where $X^n_t$ is a vector of explanatory variables including:

- The current state of the mortgage, $U^n_t$
- The features of the $n$-th loan at $t$
Deep Learning

$h_\theta(u, X_t^n)$

Borrower State Transitions

Amazon Web Services
Deep Neural Network

~300 Risk Factors:
- Loan-level
- Zip-code/state/county/national economic variables

Conditional State Transition Probabilities

Hidden Layers
Benefits of Deep Learning

- Exploits all data available (billions of monthly samples)
- Incorporates the influence of large number (~300) of risk factors
- Captures non-linear effects including interactions between factors
- Distinguishes between multiple loan states (current, 30 DPD, 60 DPD, 90 DPD, prepaid, foreclosure, REO, etc.)
- Offers superior out-of-sample predictions of borrower behavior over any future time period
### Out-of-Sample Performance

**AUC matrix (Ensemble Size 23)**

<table>
<thead>
<tr>
<th>Initial state</th>
<th>foreclosure</th>
<th>90+DQ</th>
<th>60DQ</th>
<th>30DQ</th>
<th>current</th>
</tr>
</thead>
<tbody>
<tr>
<td>foreclosure</td>
<td>0.84</td>
<td>0.72</td>
<td>0.67</td>
<td>0.71</td>
<td>0.76</td>
</tr>
<tr>
<td>90+DQ</td>
<td>0.90</td>
<td>0.82</td>
<td>0.64</td>
<td>0.69</td>
<td>0.89</td>
</tr>
<tr>
<td>60DQ</td>
<td>0.93</td>
<td>0.92</td>
<td>0.70</td>
<td>0.68</td>
<td>0.76</td>
</tr>
<tr>
<td>30DQ</td>
<td>0.69</td>
<td>0.68</td>
<td>0.70</td>
<td>0.73</td>
<td>0.83</td>
</tr>
<tr>
<td>current</td>
<td>0.66</td>
<td>0.69</td>
<td>0.81</td>
<td>0.53</td>
<td>0.74</td>
</tr>
<tr>
<td>90+DQ</td>
<td>0.77</td>
<td>0.91</td>
<td>0.91</td>
<td>0.50</td>
<td>0.76</td>
</tr>
<tr>
<td>60DQ</td>
<td>0.74</td>
<td>0.79</td>
<td>0.74</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30DQ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>current</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>90+DQ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>60DQ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30DQ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>current</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Final state**

- current
- 30DQ
- 60DQ
- 90+DQ
- foreclosure
- REO
- paidOff
Out-of-Sample Performance

(12 month ahead, portfolio size 1000)
Predictions Address Feature Volatility
## Investment Portfolio Design

### Table 4: Performance of portfolio after (out-of-sample) 12 months recorded via percent of portfolio in each state.

<table>
<thead>
<tr>
<th>State</th>
<th>NN (5)</th>
<th>LR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paid off</td>
<td>4.06</td>
<td>8.14</td>
</tr>
<tr>
<td>Current</td>
<td>93.28</td>
<td>89.09</td>
</tr>
<tr>
<td>30 days delinquent</td>
<td>1.60</td>
<td>1.54</td>
</tr>
<tr>
<td>60 days delinquent</td>
<td>0.36</td>
<td>0.36</td>
</tr>
<tr>
<td>90+ days delinquent</td>
<td>0.49</td>
<td>0.55</td>
</tr>
<tr>
<td>Foreclosure</td>
<td>0.19</td>
<td>0.30</td>
</tr>
<tr>
<td>REO</td>
<td>0.02</td>
<td>0.03</td>
</tr>
</tbody>
</table>

Select best 20,000 out of 100,000 loans according to predicted probability of being current in 12 months. Performance of portfolio after (out-of-sample) 12 months recorded via percent of portfolio in each state.
Fitted Delinquency Rate

The graph illustrates the fitted delinquency rate as a function of the zip-code level house price change since origination. Two lines represent different state unemployment rates: 6% and 12%.
Explainability

- Mortgage underwriting is heavily regulated
- Regulators insist on explainability of underwriting decisions
  
  Reason for rejection
  
  Fairness, discrimination

- We have developed statistical *significance tests* for neural networks that can be used to address these requirements
  
  Risk factor importance and ranking
<table>
<thead>
<tr>
<th>Variable</th>
<th>Gradient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Times 30 Days Delinquent in Last 12 Months</td>
<td>0.0650</td>
</tr>
<tr>
<td><strong>FICO Score</strong></td>
<td>0.0445</td>
</tr>
<tr>
<td>Number of Times 60 Days Delinquent in Last 12 Months</td>
<td>0.0334</td>
</tr>
<tr>
<td>Current Outstanding Balance</td>
<td>0.0320</td>
</tr>
<tr>
<td>Original Loan Balance</td>
<td>0.0285</td>
</tr>
<tr>
<td>Original Interest Rate</td>
<td>0.0235</td>
</tr>
<tr>
<td><strong>Zillow Zip Code Housing Price Change Since Origination</strong></td>
<td>0.0187</td>
</tr>
<tr>
<td><strong>Original Interest Rate - National Mortgage Rate</strong></td>
<td>0.0170</td>
</tr>
<tr>
<td>Number of Times 90+ Days Delinquent in Last 12 Months</td>
<td>0.0145</td>
</tr>
<tr>
<td><strong>Lagged Prime Default Rate in Same Zip Code</strong></td>
<td>0.0116</td>
</tr>
<tr>
<td>Number of Times Foreclosed in Last 12 Months</td>
<td>0.0109</td>
</tr>
<tr>
<td>Zillow zip code median house price change since origination</td>
<td>0.0108</td>
</tr>
<tr>
<td>Number of Days Delinquent</td>
<td>0.0095</td>
</tr>
<tr>
<td>Number of Times Current in Last 12 Months</td>
<td>0.0088</td>
</tr>
<tr>
<td>Time Since Origination</td>
<td>0.0087</td>
</tr>
<tr>
<td>Current Interest Rate - Original Interest Rate</td>
<td>0.0087</td>
</tr>
<tr>
<td>Lagged Prime Prepayment Rate in Same Zip Code</td>
<td>0.0074</td>
</tr>
<tr>
<td>ARM Rate Reset Frequency</td>
<td>0.0070</td>
</tr>
<tr>
<td>Total Number of Prime Mortgages in Same Zip Code</td>
<td>0.0068</td>
</tr>
<tr>
<td>Current Interest Rate - National Mortgage Rate</td>
<td>0.0065</td>
</tr>
<tr>
<td><strong>State Unemployment Rate</strong></td>
<td>0.0060</td>
</tr>
<tr>
<td>Scheduled Interest and Principle Due</td>
<td>0.0050</td>
</tr>
<tr>
<td><strong>LTV Ratio</strong></td>
<td>0.0050</td>
</tr>
<tr>
<td>Lagged Default Rate for Subprime Mortgages in Same Zip Code</td>
<td>0.0050</td>
</tr>
<tr>
<td>Original Term of the Loan</td>
<td>0.0041</td>
</tr>
<tr>
<td>:</td>
<td>:</td>
</tr>
</tbody>
</table>
Summary

• Machine intelligence can potentially transform housing finance

• We are working with the largest players in the housing finance space to realize this potential in practice

• We have also developed powerful computational algorithms that offer significant speedups for pool- and security level problems

• Applications in other lending areas: consumer loans, credit cards, student loans, auto loans, SME credit, etc.
• Deep Learning for Mortgage Risk (2016)
• Risk Analysis for Large Pools of Loans (2017)
• Large-Scale Loan Portfolio Selection (2016)
• Towards Explainable AI: Significance Tests for Neural Nets (2018)
• Machine Learning estimators for US House Prices (2019, forthcoming)
• Understanding Distressed Sale Discounts (2019, forthcoming)
• Real Estate Investment via Deep Learning (2019, forthcoming)
• Risk Premia in Mortgage-Backed Securities Markets (2019, forthcoming)