Using R in Academic Finance

Sanjiv R. Das
Professor, Santa Clara University
Department of Finance
http://algo.scu.edu/~sanjivdas/

Outline

- High-performance computing for Finance
- Modeling the optimal modification of home loans using R
- Identifying systemically risky financial institutions using R network models.
- Goal-based portfolio optimization with R
- Using R to deliver functions/models on the web, and for pedagogical purposes.

R works well with Python and C.

R/Finance 2011: Applied Finance with R

April 29 & 30, Chicago, IL, USA

home agenda travel committee 2009 Conference 2010 Conference

The **2011 presentations** are now online.

R/Finance 2012 will take place May 10 to 12, 2012!

http://www.rinfinance.com/RinFinance2010/agenda/

http://cran.r-project.org/web/views/Finance.html

Journal Of Investment Management, Vol. 7, No. 4, (2009), pp. 1–12

www.joim.com

SURVEYS AND CROSSOVERS

FINANCIAL APPLICATIONS WITH PARALLEL R

Sanjiv R. Das[a,*] *and Brian Granger*[b]

The use of statistical packages in finance has two functions. One, econometric analysis of large volumes of data, and two, programming financial models. A popular package for these purposes is R. In this article we will examine two canonical applications of parallel programming for option pricing. We use the ParallelR package developed by REvolution Computing. We price options using trees and Monte Carlo simulation. Both these approaches are commonly used for option pricing and are amenable to parallelization and grid computing. In this paper we demonstrate the application using the widely used mathematical/statistical R package.

Calling C from R: An Example of Tax-optimized Portfolio Rebalancing

THE IMPORTANCE OF REBALANCING, JANUARY 1996–DECEMBER 2009

During This Period, an Annually Rebalanced Portfolio Provided Lower Volatility and Higher Return

	Average Annual Return	*Risk* (Volatility)*
60% Russell 3000/40% Barclays Aggregate Bond: Annually Rebalanced†	10.07	11.84
60% Russell 3000/40% Barclays Aggregate Bond: Never Rebalanced†	8.76	13.12

*Standard deviation of return.

†Stocks represented by a Russell 3000 Total Stock Market Fund. Bonds represented by a Barclays Aggregate Total Bond Market Fund. (Taxes not considered.)

Calling C from R: An Example of Tax-optimized Portfolio Rebalancing

```c
#include <stdio.h>
#include <math.h>
#include <stdlib.h>
#include <time.h>
#include <string.h> /* memset */
#include <unistd.h> /* close */

#define max( a, b ) ( ((a) > (b)) ? (a) : (b) )
#define min( a, b ) ( ((a) < (b)) ? (a) : (b) )

/***** PROCESS EACH ITERATION TO GET FINAL WEALTH *****/
void ProcessWealth(double *w_in, int *n_in, int *T_in, int *ppy, int *dead_in,
           double *h_in, double *r_in, double *fl_in, double *fu_in,
           double *tau_g_in, double *tau_l_in, double *limit_in,
           double *S, double *W) {
  int n = n_in[0];
  int i, j, J, k, t;
  double Cash, LB, sumN, f, reduction, cfLoss, yr, pvLB;
  double B[n], N[n];
  double w = w_in[0];
  double wealth = W[0];
  int T = T_in[0];
  int periods_per_year = ppy[0];
  double h = h_in[0];
  double r = r_in[0];
  double f_l = fl_in[0];
  double f_u = fu_in[0];
  double tau_g = tau_g_in[0];
  double tau_l = tau_l_in[0];
  double limit = limit_in[0];
  int dead = dead_in[0];

  Cash = (1.0 - w)*wealth;          /* initialize cash balance */
  J = 1;                   /* number of bases */
  B[0] = S[0];                 /* a vector that represents stock basis */
  N[0] = w*wealth/B[0];            /* vector of number of shares for each basis */
  LB = 0.0;                /* loss balance ( > 0 for loss, < 0 for gain) */
```

R CMD SHLIB tax.c

```
dyn.load("tax.so")

#SHELL FUNCTION TO CALL C CODE
PW = function(w_in,n_in,T_in,ppy,dead_in,h_in,r_in,fl_in,fu_in,tau_g_in,tau_l_in,limit_in,S,W) {
        pw = .C("ProcessWealth",w_in,n_in,T_in,ppy,dead_in,h_in,r_in,fl_in,fu_in,tau_g_in,tau_l_in,limit_in,S,W)
}

#FUNCTION TO RETURN EXPECTED UTILITY (TWO PARAMETERS {w,delta})
EU2 = function(p,n_in,T_in,ppy,dead_in,h_in,r_in,tau_g_in,tau_l_in,limit_in,Smat,W,alpha) {
    w_in = max(0,min(1,as.double(p[1]))); fl_in = max(0,as.double(w_in - p[2])); fu_in = min(1,as.double(w_in + p[2]));
    m = length(Smat[,1])
    finalW = matrix(0,m,1)
    for (j in seq(1,m)) {
        Srow = as.double(Smat[j,])
        finalW[j] = PW(w_in,n_in,T_in,ppy,dead_in,h_in,r_in,fl_in,fu_in,tau_g_in,tau_l_in,limit_in,Srow,W)[[14]][1]
    }
    eu = mean((finalW^(1-alpha))/(1-alpha))
    eu = -1000000000000.0*eu   #To be used for maximization if a minimizer is used, else comment it out

}

####### INITIALIZE VARIABLES THAT ARE PASSED TO FUNCTION #######
w_in = as.double(0.5)           #   Initial portfolio weight
n_in = as.integer(480)          #   Total number of periods
T_in = as.integer(40)           #   Number of years
ppy = as.integer(12)            #   Number of rebalances per year (periods per year)
dead_in = as.integer(1)         #   dead = 1 means dead at time T, dead = 0 means alive at time T
h_in = as.double(1.0/ppy)       #   Length of each period
r_in = as.double(0.03)          #   Annual risk-free rate
fl_in = as.double(0.00)         #   Lower bound
fu_in = as.double(1.00)         #   Upper bound
tau_g_in = as.double(0.15)      #   Capital gain tax rate
```

Topic 1

MODIFYING HOME LOANS WITH R MODELS

THE PRINCIPAL PRINCIPLE: Optimal Modification of Distressed Home Loans
(Why Lenders should Forgive, not Foresake Mortgages)

STRATEGIC LOAN MODIFICATION: An Options based response to strategic default
(joint work with Ray Meadows)

CitiMortgage CEO Sanjiv Das helps people keep their homes

By Stephanie Armour, USA TODAY
April 27, 2009

There is very little in Sanjiv Das' uncluttered office.

Just snapshots of his wife and his 20-year-old daughter, Natasha. No pictures of his favorite sports: golf and cricket.

Das has moved around the world, run credit card acquisitions for American Express in India, and handled the mortgage business for Citibank in Sydney. But none of it compares to what he's doing now.

A tale of two Das: Citi CEO, academic and mortgages

Al Yoon
NEW YORK
Wed Feb 24, 2010 [illegible]

NEW YORK (Reuters) - Sanjiv Ranjan Das, a professor at California's Santa Clara University, last fall attacked the problem of "underwater" mortgages often cited as an Achilles' heel to the U.S. housing market.

HOUSING MARKET

He had a special fan: Sanjiv Das, the top executive at CitiMortgage, the nation's fourth-largest home loan lender and servicer of $723 billion in mortgages.

HAMP UPDATE

June 4, 2010

Announcing HAMP Principal Reduction Alternative

Yesterday, June 3, 2010, Supplemental Directive 10-05: Modification of Loans with Principal Reduction Alternative, was issued offering mortgage relief to eligible homeowners whose homes are worth significantly less than the remaining amounts owed under their first lien mortgage loans. The Principal Reduction Alternative (PRA) guidance applies to non-GSE loans eligible for the Home Affordable Modification Program (HAMP) only.

Principal Reduction Alternative (PRA)
With this new guidance, servicers are required to evaluate all HAMP-eligible loans with a mark-to-market loan-to-value (MTMLTV) greater than 115% to determine if a principal reduction is beneficial. If the evaluation shows the net present value (NPV) for a HAMP modification using PRA is positive, servicers are encouraged to offer the principal reduction to the borrower. An updated NPV model reflecting principal reduction will be available to use for this evaluation. Additional details are as follows:

- **Effective Date** -- The PRA Effective Date (i.e., the date the principal reduction evaluation is required) will be either October 1, 2010, or the date of the HAMP NPV Model 4.0 release (whichever is later). However, servicers may immediately offer PRA for HAMP-eligible modifications as long as the reduction follows all PRA requirements.
- **Application** -- PRA is earned over a three-year period and is initially treated as a PRA Forbearance. Each year (for three years) that the borrower is in good standing on the anniversary of their trial period effective date, one-third of the original PRA forbearance amount will be reduced. This reduced amount will be applied to their unpaid principal balance.
- **Second Lien** -- Servicers participating in the Second Lien Program (2MP) will be required to provide a principal reduction on the borrower's second mortgage in proportion to any principal reduction offered on the borrower's first mortgage.
- **Investor Incentive** -- Investors will receive an incentive based on loan delinquency, LTV ratio, and the amount of the principal reduction. Note: Guidance on principal reduction and related investor incentives will be forthcoming for loans in active HAMP Trial Period Plans or that were permanently modified prior to June 3, 2010 (i.e., the SD 10-05 effective date).

Game theoretic problem:

Lender determines the loan modification that maximizes value of loan given that the borrower will act strategically in his best interest.

Model

Home value

$$\frac{dV(t)}{V(t)} = (r - \delta)V(t)\ dt + \sigma_1\ dZ_1(t) + \sigma_2\ dZ_2(t)$$

HJM

$$df(t,T) = \alpha(t,T)\ dt + \beta(t,T)\ dZ_1(t), \quad \forall T.$$

Correlation

$$Corr(dV/V, df) = \frac{\sigma_1}{\sqrt{\sigma_1^2+\sigma_2^2}}$$

Discrete-time Implementation

$$V(t+1) = \begin{cases} V(t)\exp\left(+\sigma_1\sqrt{h}+\sigma_2\sqrt{h}\right), & \text{w/prob } q/2 \\ V(t)\exp\left(+\sigma_1\sqrt{h}-\sigma_2\sqrt{h}\right), & \text{w/prob } (1-q)/2 \\ V(t)\exp\left(-\sigma_1\sqrt{h}+\sigma_2\sqrt{h}\right), & \text{w/prob } q/2 \\ V(t)\exp\left(-\sigma_1\sqrt{h}-\sigma_2\sqrt{h}\right), & \text{w/prob } (1-q)/2 \end{cases}$$

$$f(t+1,T) = \begin{cases} f(t,T)+\alpha(t,T)\,h+\beta(t,T)\sqrt{h}, & \text{w/prob } \frac{1}{2} \\ \\ f(t,T)+\alpha(t,T)\,h-\beta(t,T)\sqrt{h}, & \text{w/prob } \frac{1}{2} \end{cases}$$

$$\sum_{t=1}^{T}\alpha(t,T) \;=\; \frac{1}{h^2}\ln\left[\cosh\left(\sum_{t=1}^{T}\beta(t,T)\,h^{\frac{3}{2}}\right)\right], \quad \forall T$$

Martingale system

$$
\begin{aligned}
V(t)\, e^{(r(t)-\delta)h} &= E[V(t+1)] \\
&= V(t) \exp\left(+\sigma_1\sqrt{h} + \sigma_2\sqrt{h}\right) \times q/2 \\
&\quad + V(t) \exp\left(+\sigma_1\sqrt{h} - \sigma_2\sqrt{h}\right) \times (1-q)/2 \\
&\quad + V(t) \exp\left(-\sigma_1\sqrt{h} + \sigma_2\sqrt{h}\right) \times q/2 \\
&\quad + V(t) \exp\left(-\sigma_1\sqrt{h} - \sigma_2\sqrt{h}\right) \times (1-q)/2
\end{aligned}
$$

Risk-neutral Probabilities

$$q = \frac{2\ e^{(r(t)-\delta)h} - (u_2 + u_4)}{u_1 + u_3 - u_2 - u_4}$$

$$\begin{aligned}
u_1 &= \exp\left(+\sigma_1\sqrt{h} + \sigma_2\sqrt{h}\right) \\
u_2 &= \exp\left(+\sigma_1\sqrt{h} - \sigma_2\sqrt{h}\right) \\
u_3 &= \exp\left(-\sigma_1\sqrt{h} + \sigma_2\sqrt{h}\right) \\
u_4 &= \exp\left(-\sigma_1\sqrt{h} - \sigma_2\sqrt{h}\right)
\end{aligned}$$

Modeling the Mortgage

Loan balance $$L(t+1) = L(t)\left[(1+i) - \frac{i}{1-(1+i)^{-(N-t)}}\right]$$

Default put $$P(T) = \max(0, S - V(T) - K_R)$$

Lender's value on default $$B^l(T) = \phi \cdot V(T)$$

Borrower's liability $$B^b(T)$$

Deadweight cost of foreclosure $$(1-\phi)\ V(T)$$

Refinancing option $$B^b(t) > L(t) + S?$$

"Iso-Service" Surface

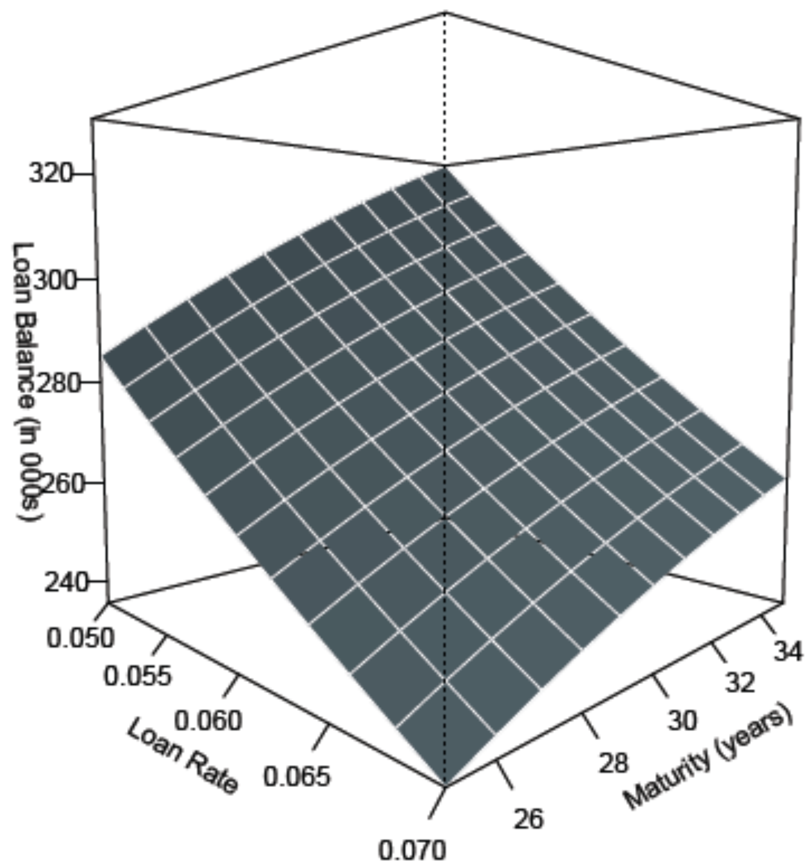
$A \leq A_{max}$ → choose → $\{T, L_t, r_L\}$

Loan balance = \$300,000
Home value = \$250,000

Remaining maturity = 25 years
A = \$1,933 per month

A_{max} = \$20,000 per year
(\$1,667 per month)

$$L_0 = \frac{A_{max}}{m}\left[\frac{1-(1+r_L/m)^{-N}}{r_L/m}\right]$$

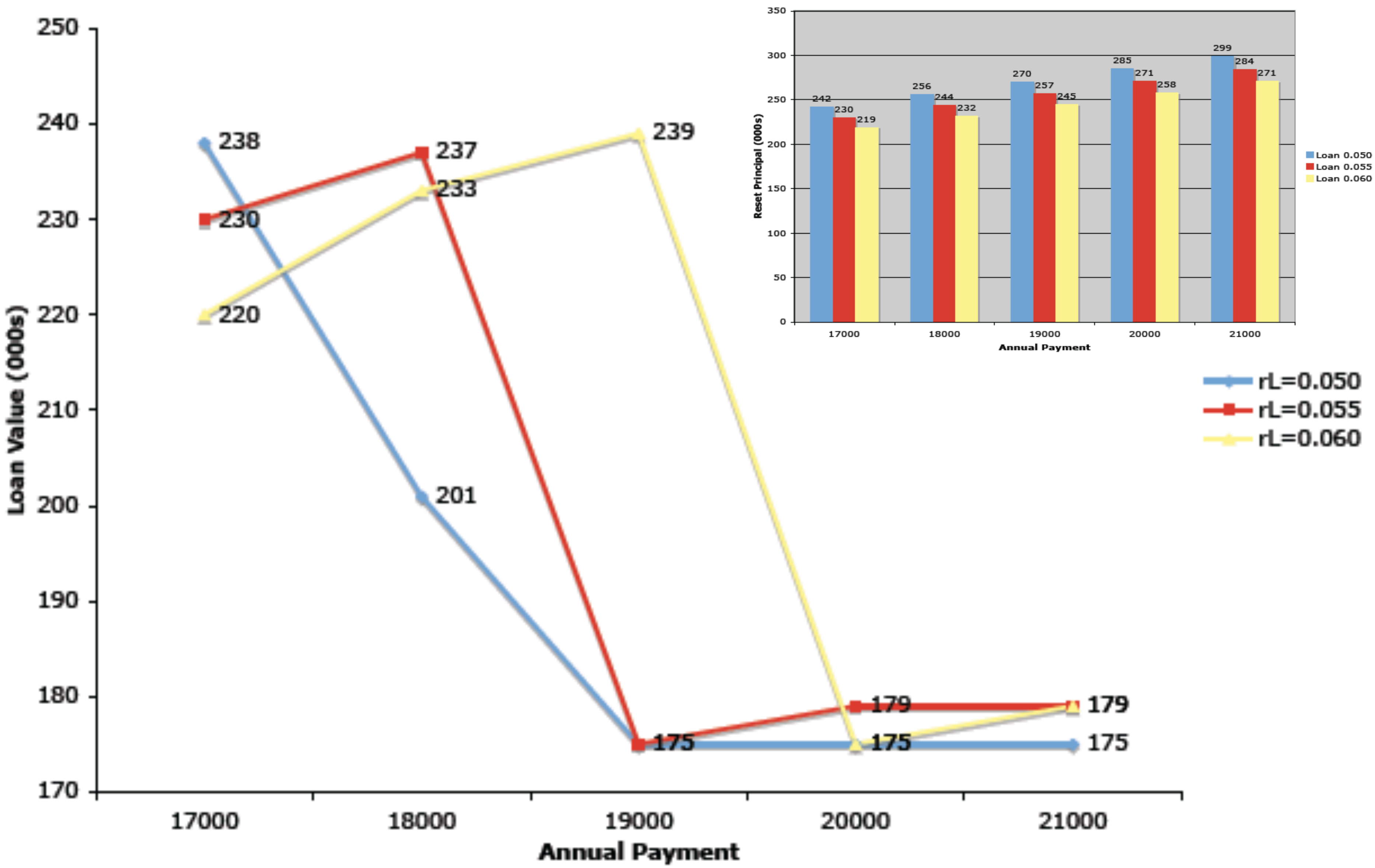


Some R

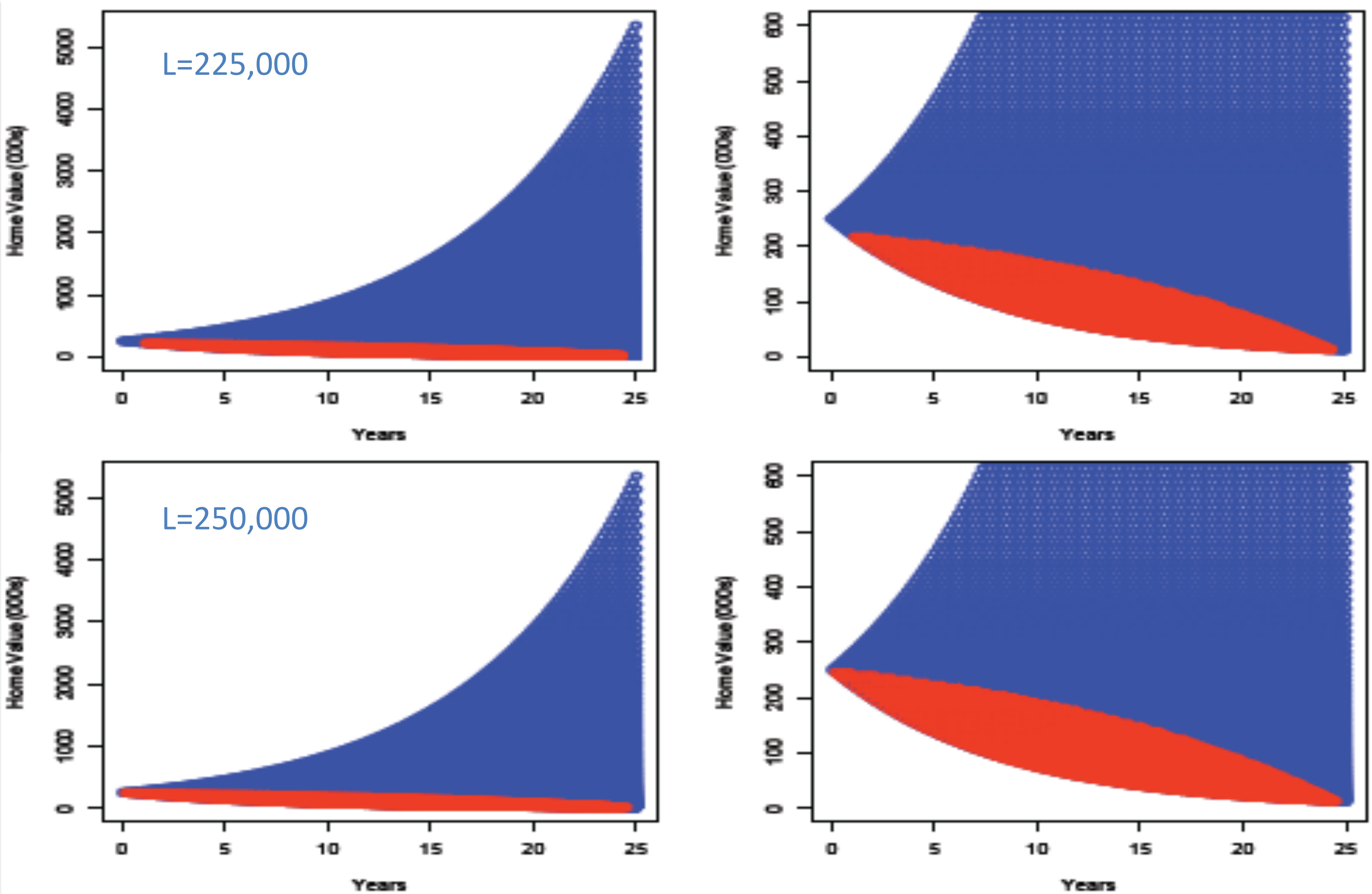
homeloan.R

```r
# Generate the forward rates and spot rates lattices
# Note that in R, the indexes run backward, i.e., (j,i,t), its weird but works
F = array(0,dim=c(maxm,maxm,maxm))
spotRf = array(0,dim=c(maxm,maxm))
F[,1,1] = fwdr
spotRf[1,1] = fwdr[1]
for (t in 1:(maxm-1)) {
   for (i in 1:t) {
      f0 = F[1:(maxm-t+1),i,t]
      sig0 = fsig[t:maxm]
      u = hjm(f0,sig0,h)
      lenu = length(u[,1])
      F[(1:lenu),i,t+1] = u[,1]
      F[(1:lenu),i+1,t+1] = u[,2]
      spotRf[t+1,i] = F[1,i,t+1]
      spotRf[t+1,i+1] = F[1,i+1,t+1]
      #Inspection of spotRf shows that rates percolate down the tree, not across.
   }
}


# Create the stock price tree, spor rates (R)
# Stock prices percolate down the tree, not across.
R = array(0,dim=c(maxm,maxm,maxm))
R[1,1,1] = spotRf[1,1]
V = array(0,dim=c(maxm+1,maxm+1,maxm+1))
u1 = exp(sig1*sqrt(h)+sig2*sqrt(h));
u2 = exp(sig1*sqrt(h)-sig2*sqrt(h));
u3 = exp(-sig1*sqrt(h)+sig2*sqrt(h));
u4 = exp(-sig1*sqrt(h)-sig2*sqrt(h));
shift_V = matrix(c(u1,u3,u2,u4),2,2)
V[1,1,1] = v0;
for (t in 2:(maxm+1)) {
    for (i in 1:(t-1)) {
        for (j in 1:(t-1)) {
            V[j:(j+1),i:(i+1),t] = V[j,i,t-1]*shift_V
            if (t<=maxm) { R[j,i:(i+1),t] = spotRf[t,j] }
        }
        if (t<=maxm) { R[t,i:(i+1),t] = spotRf[t,t] }
    }
}
```

Values of Iso-Service Loans



Default Put Exercise Region



Cure risk and Re-default Risk

The risk of unnecessary relief, i.e., the borrower would not have ultimately defaulted.

Providing futile relief, leading to ultimate default anyway.

$$\text{Loan Value} = \mathcal{B}(A) \cdot \left[1 - N\left(\frac{A - \mu_I}{\sigma_I}\right)\right] + \phi V_0 \cdot N\left(\frac{A - \mu_I}{\sigma_I}\right)$$

Value of loan accounting for willingness to pay

A: borrower income available for housing service, with mean μ and std. dev σ.

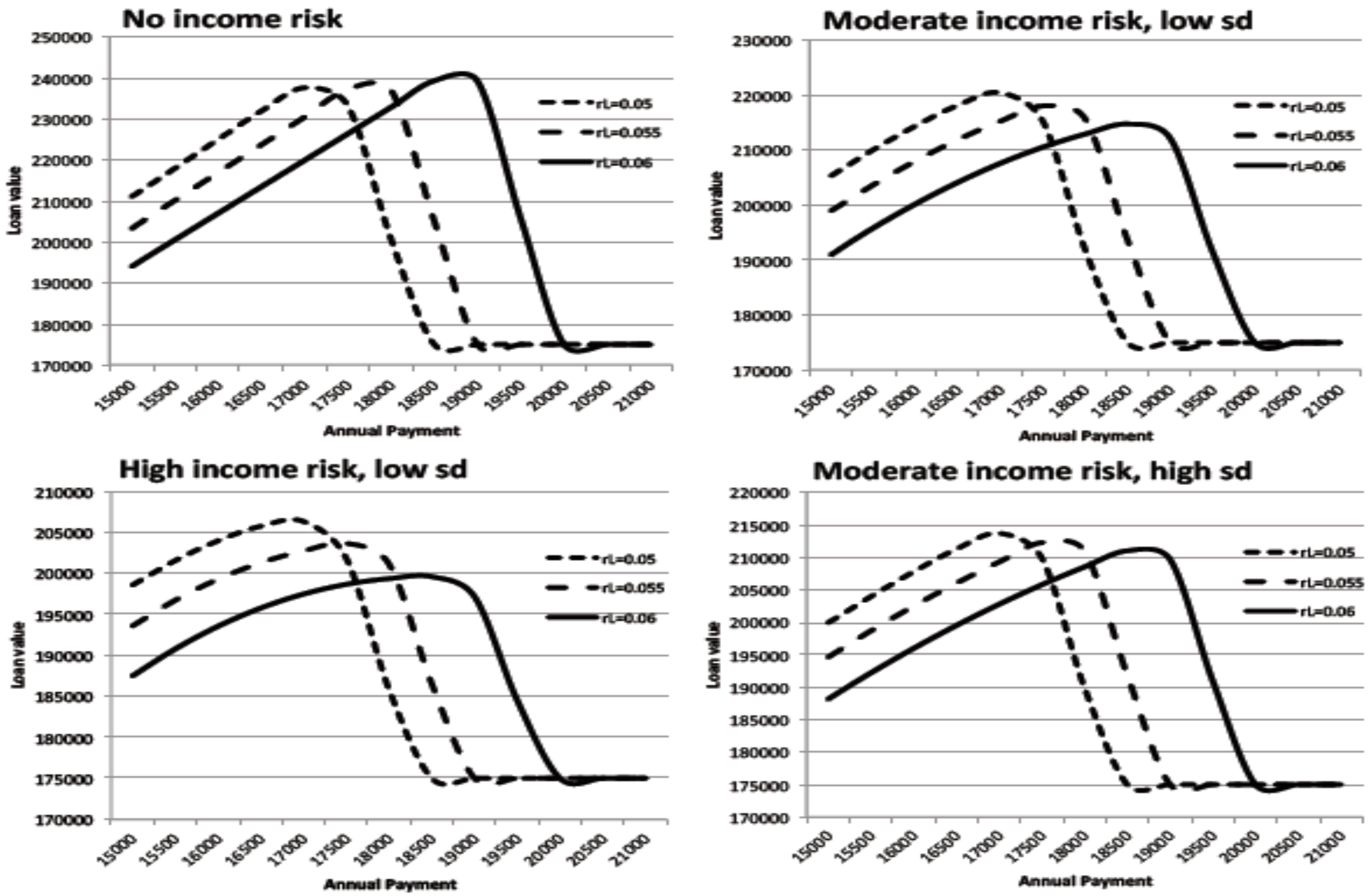


Figure 3: Loan values for differing income risk. We vary the annual payment from \$15,000 to \$21,000. For each of the these payments, we vary the loan rate in the set $r_L = \{5.0\%, 5.5\%, 6.0\%\}$. The parameters used for this graph are: home value volatility parameters $\sigma_1 = 0.02$ and $\sigma_2 = 0.03$, service flow level $\delta = 0.01$, interest rate volatility per annum $\beta = 0.0050$ (i.e., 50 bps), time step $h = 1/4$, loan rate $r_L = 0.06$, relocation costs $K_R = 0$ (increased in a later example), foreclosure recovery rate $\phi = 0.7$, loan maturity $T = 25$ years, and a flat forward rate curve at 5%. The mean and standard deviation of the ability to pay (income risk) is varied across the plots: (a) top left plot, no risk; (b) top right, mean ability to pay \$20,000 per year, sd=\$5,000, moderate risk; (c) high income risk, mean=\$17000, sd=\$5000; (d) moderate income risk, with high range, mean=\$20,000, sd=\$10,000.

Logit: Explaining Re-default

2007	< 450		450-620		620-820		820-1120		> 1120	
	Monthly Payment Amount after Modification									
	Est.	χ^2	Est.	χ^2	Est.	χ^2	Est.	χ^2	Est.	χ^2
Intercept	-2.91090	9.935	-3.23960	14.896	-1.75030	5.092	-2.71130	9.844	-1.47930	5.899
ΔRate	-0.11310	0.796	-0.28570	4.212	-0.31070	3.931	-0.33240	5.760	-0.36810	8.224
ΔTerm	-0.00640	5.427	-0.00133	0.220	0.00114	0.079	0.00016	0.002	-0.00038	0.008
ΔPrincipal	-0.00017	5.323	-0.00007	2.303	-0.00007	5.633	-0.00004	3.567	-0.00004	7.157
LTV	0.02270	6.143	0.02530	8.212	0.00163	0.040	0.02200	6.358	0.00141	0.043
Debt Ratio	-0.02110	2.021	-0.01060	0.802	-0.00291	0.056	-0.01790	3.182	-0.01150	1.357
PPC	-0.01000	0.421	-0.01340	1.888	-0.00324	0.033	-0.00463	0.106	0.00077	0.004
Re-Default	43		59		66		89		93	
No Default	230		330		382		459		545	
Wald Stat	18.9061	0.0043	18.8779	0.0044	15.1021	0.0195	23.4206	0.0007	20.8749	0.0019

2008	< 450		450-620		620-820		820-1120		> 1120	
	Monthly Payment Amount after Modification									
	Est.	χ^2	Est.	χ^2	Est.	χ^2	Est.	χ^2	Est.	χ^2
Intercept	-0.66310	3.595	-1.77870	18.532	-2.47180	39.303	-1.77290	22.438	-1.89300	39.145
ΔRate	0.13490	7.877	-0.16530	4.650	-0.01360	0.037	-0.30900	14.547	-0.14070	4.826
ΔTerm	0.00077	0.408	0.00028	0.042	0.00024	0.038	-0.00138	1.399	0.00035	0.143
ΔPrincipal	-0.00005	2.328	-0.00008	18.808	-0.00006	21.597	-0.00007	47.587	-0.00004	41.344
LTV	-0.00267	0.675	0.00806	4.207	0.01370	13.809	0.00653	3.627	0.00859	12.609
Debt Ratio	-0.01710	5.000	-0.00693	1.027	-0.00356	0.344	-0.00025	0.002	-0.00113	0.068
PPC	-0.02540	12.026	-0.00572	0.453	-0.01130	1.826	0.00540	0.417	-0.00754	1.286
Re-Default	177		205		286		375		522	
No Default	747		819		1000		1071		1564	
Wald Stat	32.3832	< 0.0001	42.1224	< 0.0001	46.1375	< 0.0001	83.5345	< 0.0001	81.2047	< 0.0001

Reduced-Form Analysis of SAMs

$$dH_t = \mu H_t \, dt + \sigma H_t \, dZ_t$$

Home values

Normalize initial home value to 1. The option to default is ITM when ($H > L$).

There is a home value D at which the borrower will default. D is a "default level" or default exercise barrier.

D is a function of the lender share θ, we write it as $D(L, \theta)$.

D increases in L and in θ.

Foreclosure recovery as a fraction of H is ϕ.

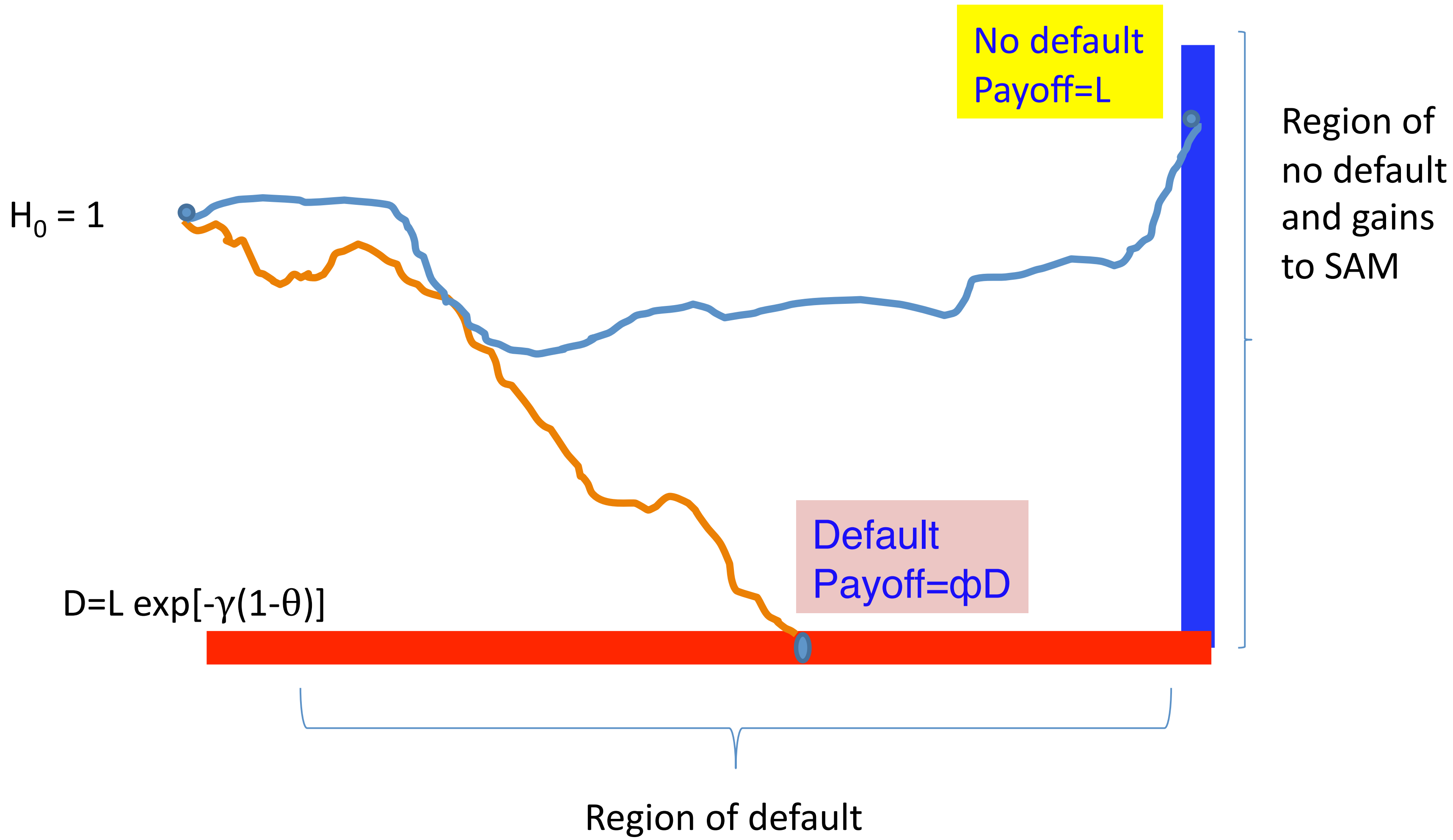
Default Barrier and Lender Share

$$D = L \exp[-\gamma(1-\theta)]$$

(1) The greater the willingness to pay (γ), the lower is the default level of home value D.
(2) When $\gamma = \infty$, the willingness to pay is infinite, the default level $D = 0$. The borrower never defaults unless the home value goes to zero.
(3) When $\gamma = 0$, there is no willingness to pay and the default level is $D = L$, i.e., the borrower defaults the moment the home value drops infintesimally below LTV at the time zero.

(1) The greater the lender's share (θ), the higher is the default level of home value D. The likelihood of default is therefore greater.
(2) When the lender share $\theta = 0$, the default level is $Le^{-\gamma}$.
(3) When $\theta = 1$, the default level is $D = L$. The borrower defaults the moment there is negative equity.

Barrier Model Intuition



A Barrier Option Decomposition

Non-default component

$$Le^{-rT}\int_{D(L,\theta)}^{\infty} p(H_T|H_t > D, \forall t < T)\, dH_T$$

where $p(H_T|H_t > D, \forall t < T)$ is the density of the terminal home value conditional on no interim default.

Default component

$$\phi D \int_0^T e^{-rt} f(t;D)\, dt$$

where $f(t;D)$ is the first-passage time density for $H_t = D$.

Shared Appreciation component

$$e^{-rT}\int_K^{\infty} (H_T - K)\, p(H_T|H_t > D, \forall t < T)\, dH_T$$

PDE

$$\frac{\partial F}{\partial H}[\mu - \lambda\sigma]H + \frac{1}{2}\frac{\partial^2 F}{\partial H^2}\sigma^2 H^2 + \frac{\partial F}{\partial t} = rF$$

The Closed-Form Solution

$$\text{LOANVAL} \equiv V(H, L, K, r, T, \phi, \theta, \mu, \lambda, \sigma, \gamma)$$

$$\begin{aligned}
&= Le^{-rT}\left[N(d_2') - (D/H)^{2(R/\sigma^2)-1} \cdot N(d_{2b}')\right] \\
&\quad + \phi D\left[(D/H)^{b_1} \cdot N(a_1) + (D/H)^{b_2} \cdot N(a_2)\right] \\
&\quad + \theta\left[C_{SAM}(H, K) - D^{2(R/\sigma^2)-1} \cdot C_{SAM}(D^2/H, K)\right]
\end{aligned}$$

$$D = L\exp[-\gamma(1-\theta)]$$

$$d_2' = \frac{\ln(H/D) + (R - 0.5\sigma^2)T}{\sigma\sqrt{T}}$$

$$d_{2b}' = \frac{\ln(D/H) + (R - 0.5\sigma^2)T}{\sigma\sqrt{T}}$$

$$a_1 = \frac{\ln(D/H) + \sqrt{2r\sigma^2 + (R - 0.5\sigma^2)^2} \cdot T}{\sigma\sqrt{T}}$$

$$a_2 = \frac{\ln(D/H) - \sqrt{2r\sigma^2 + (R - 0.5\sigma^2)^2} \cdot T}{\sigma\sqrt{T}}$$

$$b_1 = \frac{(R - 0.5\sigma^2) + \sqrt{2r\sigma^2 + (R - 0.5\sigma^2)^2}}{\sigma^2}$$

$$b_2 = \frac{(R - 0.5\sigma^2) - \sqrt{2r\sigma^2 + (R - 0.5\sigma^2)^2}}{\sigma^2}$$

$$C_{SAM}(x, y) = xe^{-(r-R)T}N(d_1') - ye^{-rT}N(d_1' - \sigma\sqrt{T})$$

$$d_1' = \frac{\ln(x/y) + (R + 0.5\sigma^2)T}{\sigma\sqrt{T}}$$

SAM or not?

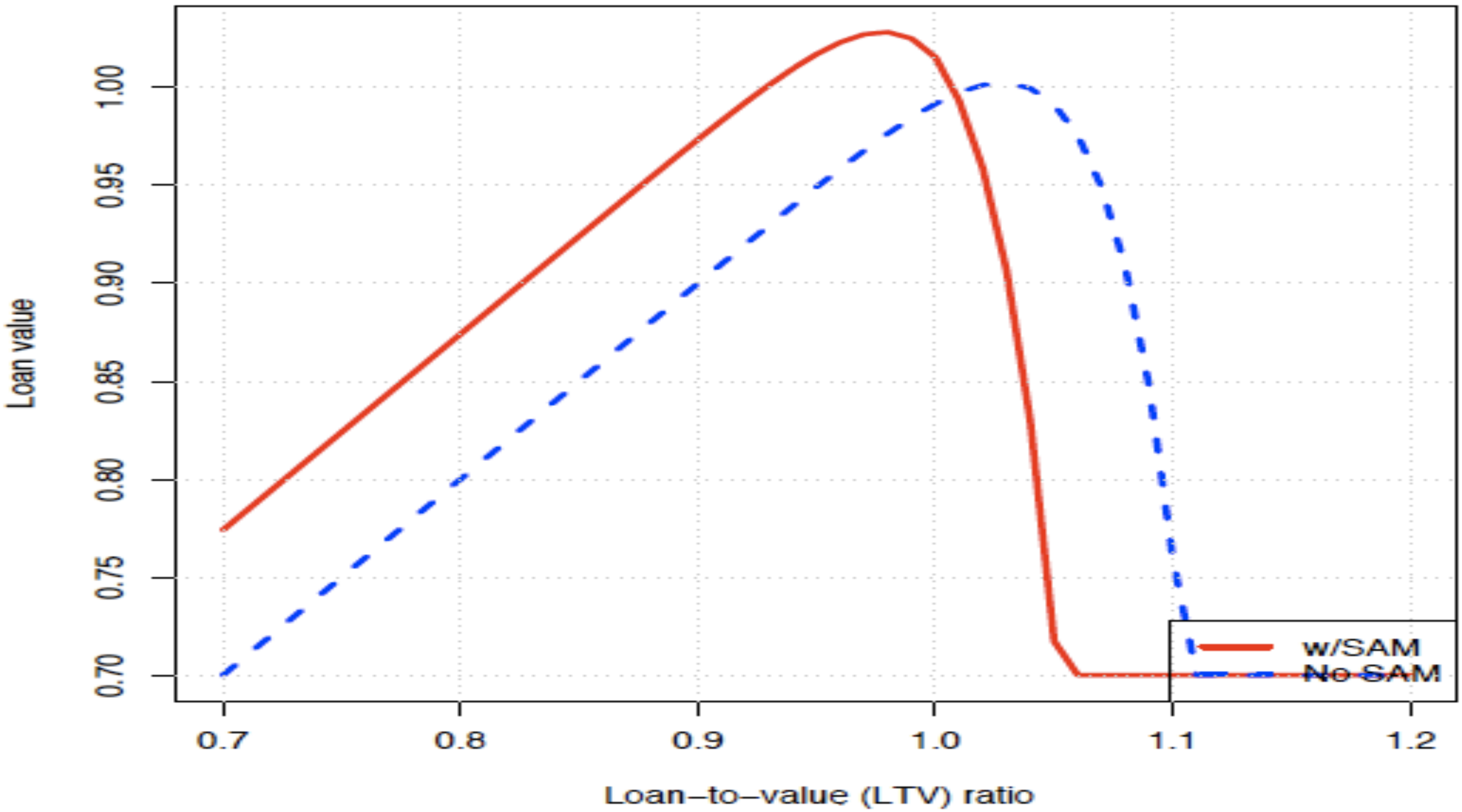


Fig. 2. Loan value as LTV is varied for loans with and without appreciation sharing. The parameters for the plot are as follows: willingness to pay coefficient $\gamma = 0.1$, home price volatility $\sigma = 0.04$, foreclosure fraction $\phi = 0.7$, risk-free rate $r = 0.02$, the house value growth rate $\mu = 0.04$, price of risk $\lambda = 0.25$, and the horizon of the model $T = 5$ years. The appreciation share fraction is $\theta = 0.50$ for the case when a SAM is applied, and $\theta = 0$ when there is no share appreciation.

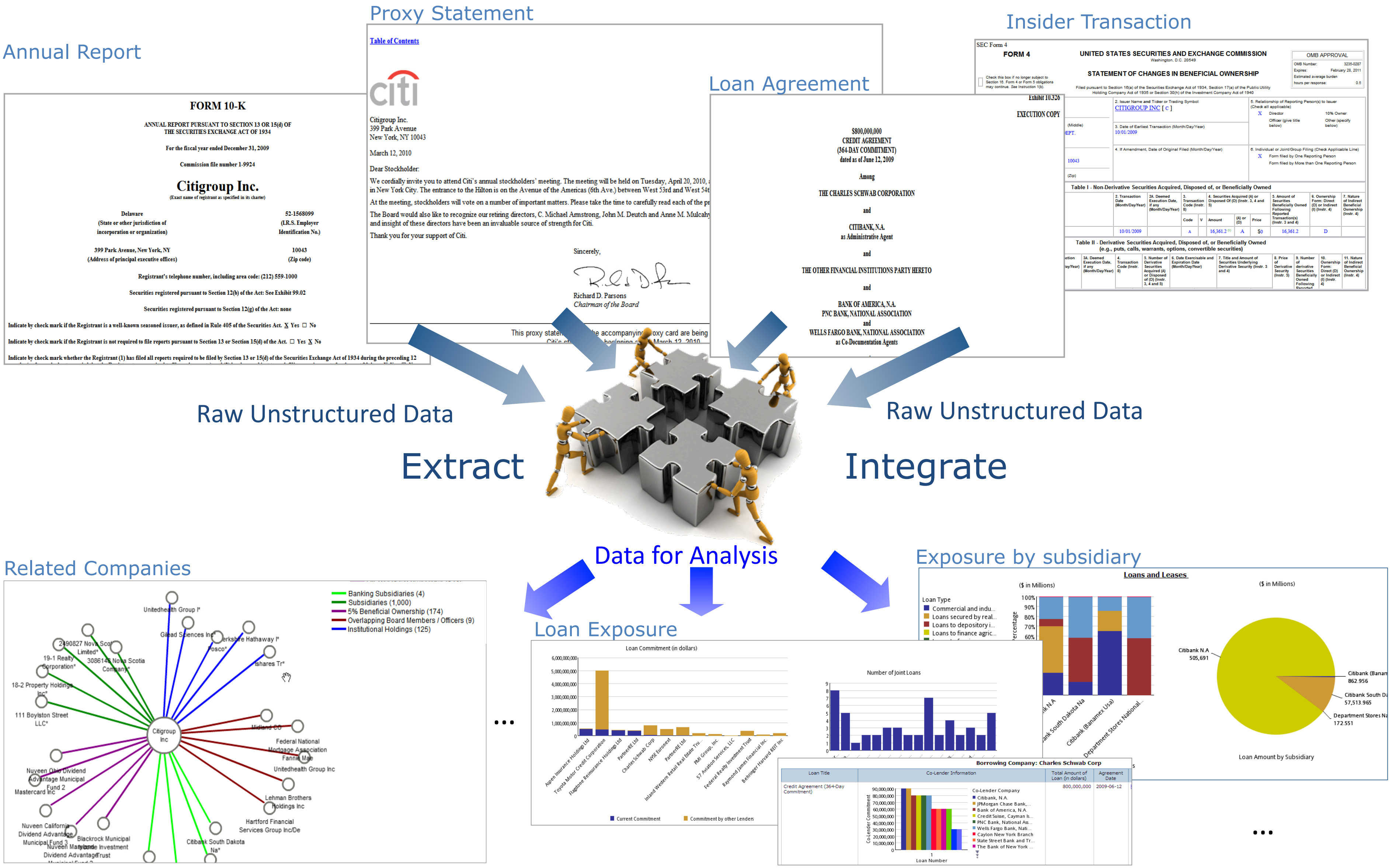
Topic 2

MANAGING SYSTEMIC RISK BY ANALYZING NETWORKS USING R

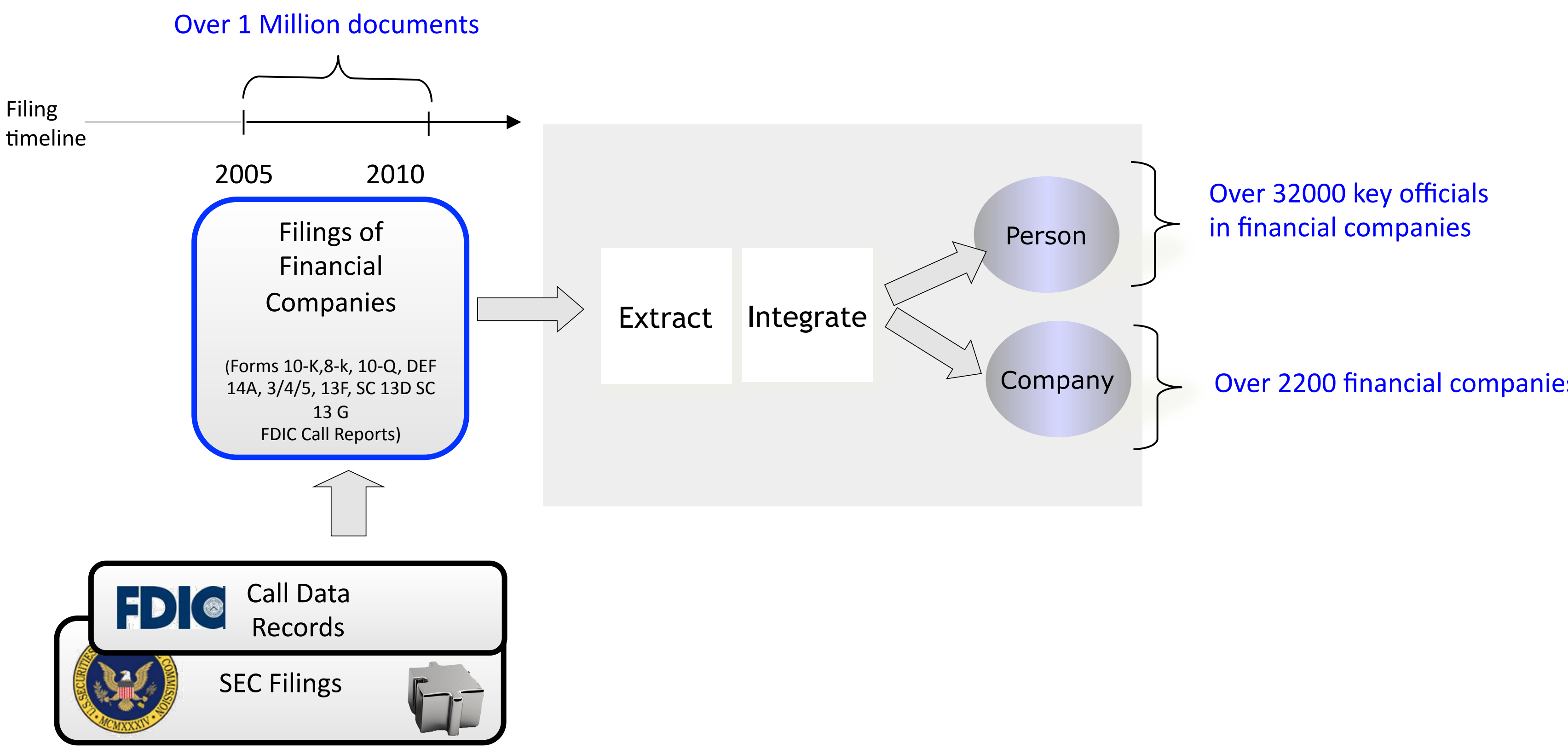
THE MIDAS PROJECT @IBM

Paper: "Unleashing the Power of Public Data for Financial Risk Measurement, Regulation, and Governance" (with Mauricio A. Hernandez, Howard Ho, Georgia Koutrika, Rajasekar Krishnamurthy, Lucian Popa, Ioana R. Stanoi, Shivakumar Vaithyanathan) IBM Almaden)

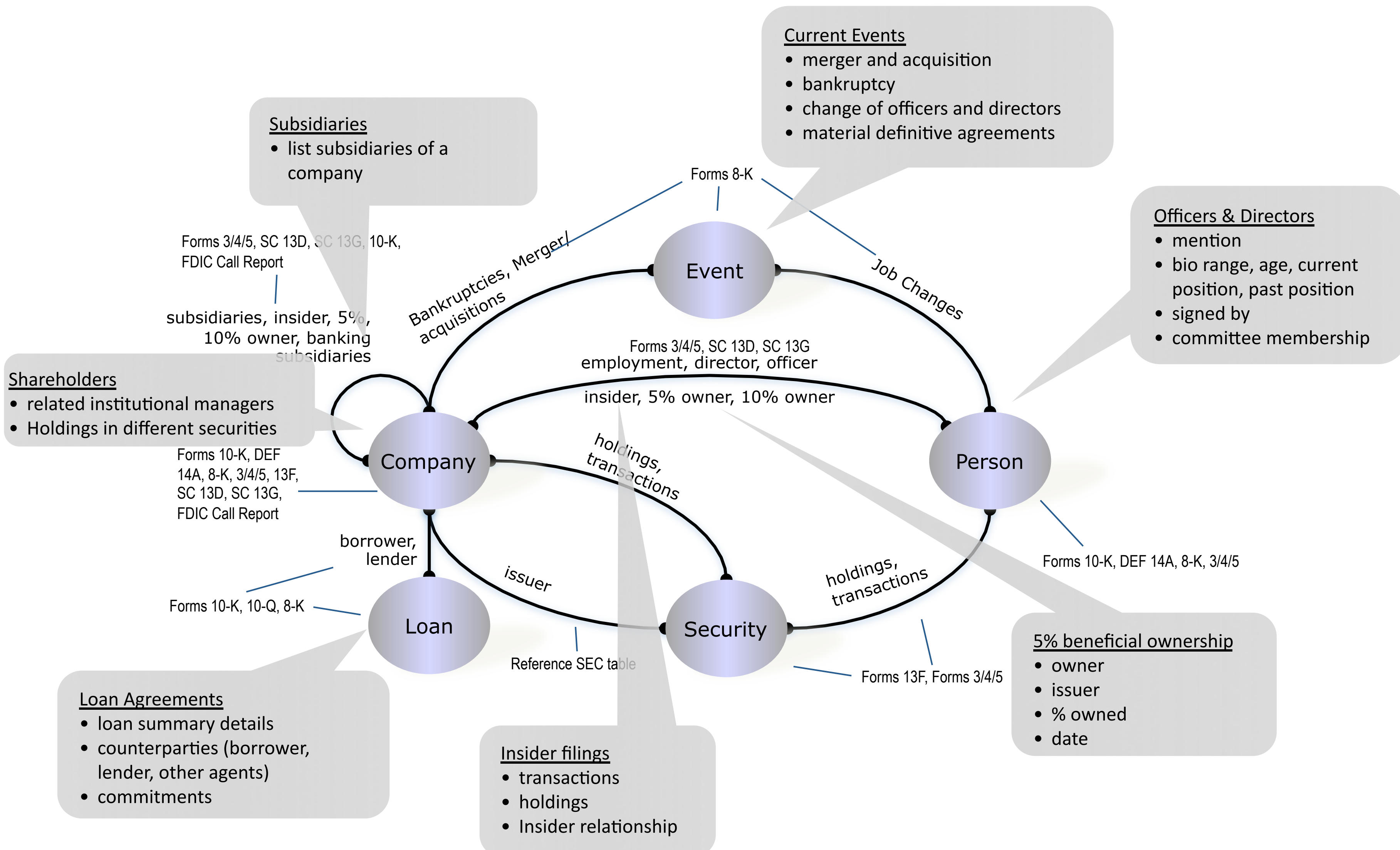
Midas Financial Insights



Example of Midas Financial Insights



Midas provides Analytical Insights into company relationships by exposing information concepts and relationships within extracted concepts



Systemic Analysis

Systemic Analysis

- Definition: the measurement and analysis of relationships across entities with a view to understanding the impact of these relationships on the system as a whole.
- Challenge: requires most or all of the data in the system; therefore, high-quality information extraction and integration is critical.

Systemic Risk

- Current approaches: use stock return correlations (indirect). [Acharya, et al 2010; Adrian and Brunnermeier 2009; Billio, Getmansky, Lo 2010; Kritzman, Li, Page, Rigobon 2010]
- Midas: uses semi-structured archival data from SEC and FDIC to construct a co-lending network; network analysis is then used to determine which banks pose the greatest risk to the system.

Co-lending Network

- Definition: a network based on links between banks that lend together.
- Loans used are *not* overnight loans. We look at longer-term lending relationships.
- Lending adjacency matrix:
- Undirected graph, i.e., symmetric
- Total lending impact for each bank:

$$\mathbf{L} \equiv \{L_{ij}\}, i, j = 1 \ldots N$$

$$\mathbf{L} \in R^{N \times N}$$

$$x_i, i = 1 \ldots N.$$

Centrality

- Influence relations are circular:

$$x_i = \sum_{j=1}^{N} L_{ij} x_j, \forall i.$$

$$\mathbf{x} = \mathbf{L} \cdot \mathbf{x}, \text{ where } \mathbf{x} = [x_1, x_2, \ldots, x_N]' \in R^{N \times 1}$$

- Pre-multiply by scalar to get an eigensystem:

$$\lambda \ \mathbf{x} = \mathbf{L} \cdot \mathbf{x}$$

- Principal eigenvector of this system gives the "centrality" score for a bank.

- This score is a measure of the systemic risk of a bank.

Data

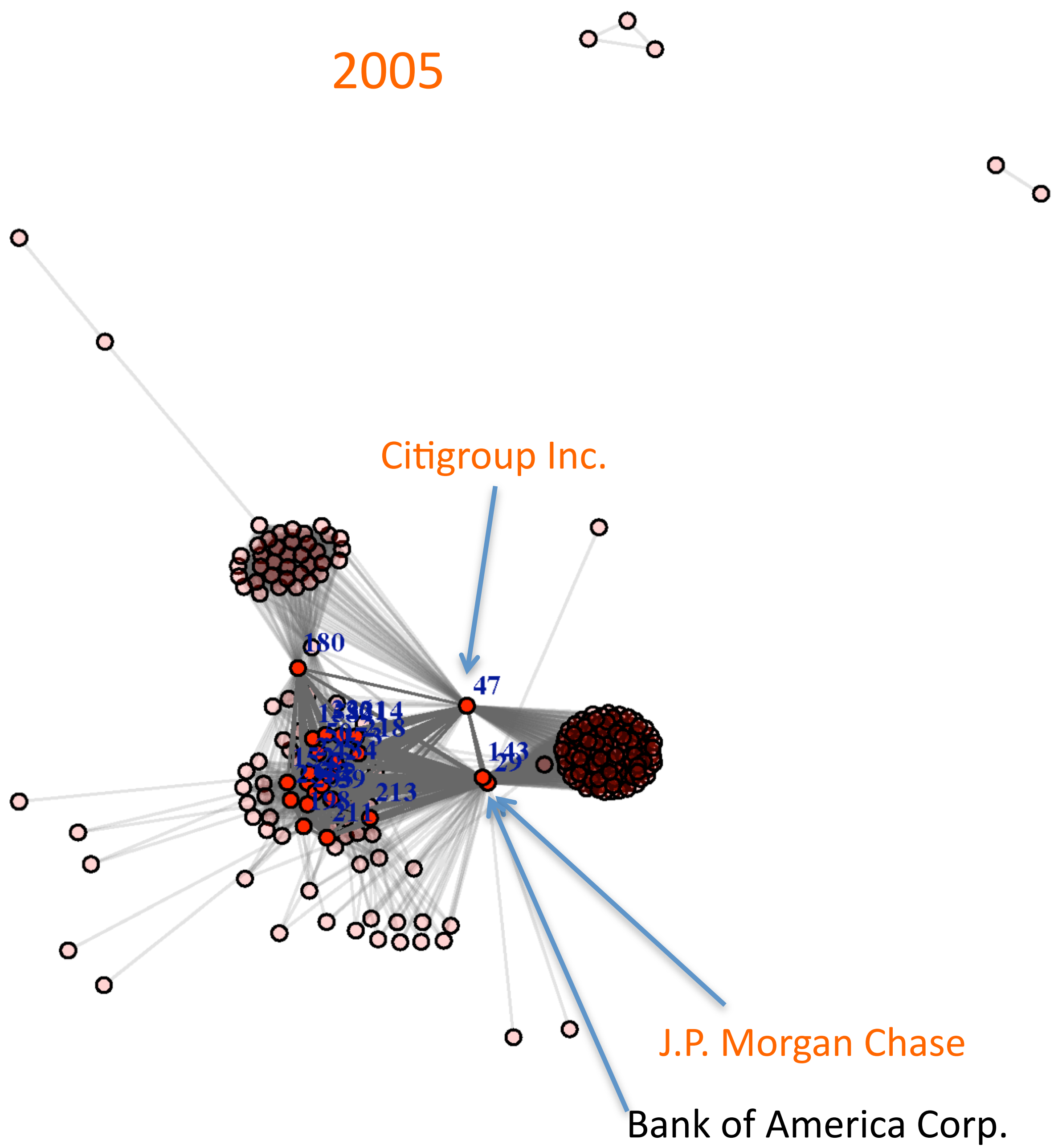
- Five years: 2005—2009.
- Loans between FIs only.
- Filings made with the SEC.
- No overnight loans.
- Example: 364-day bridge loans, longer-term credit arrangement, Libor notes, etc.
- Remove all edge weights < 2 to remove banks that are minimally active. Remove all nodes with no edges. (This is a choice for the regulator.)

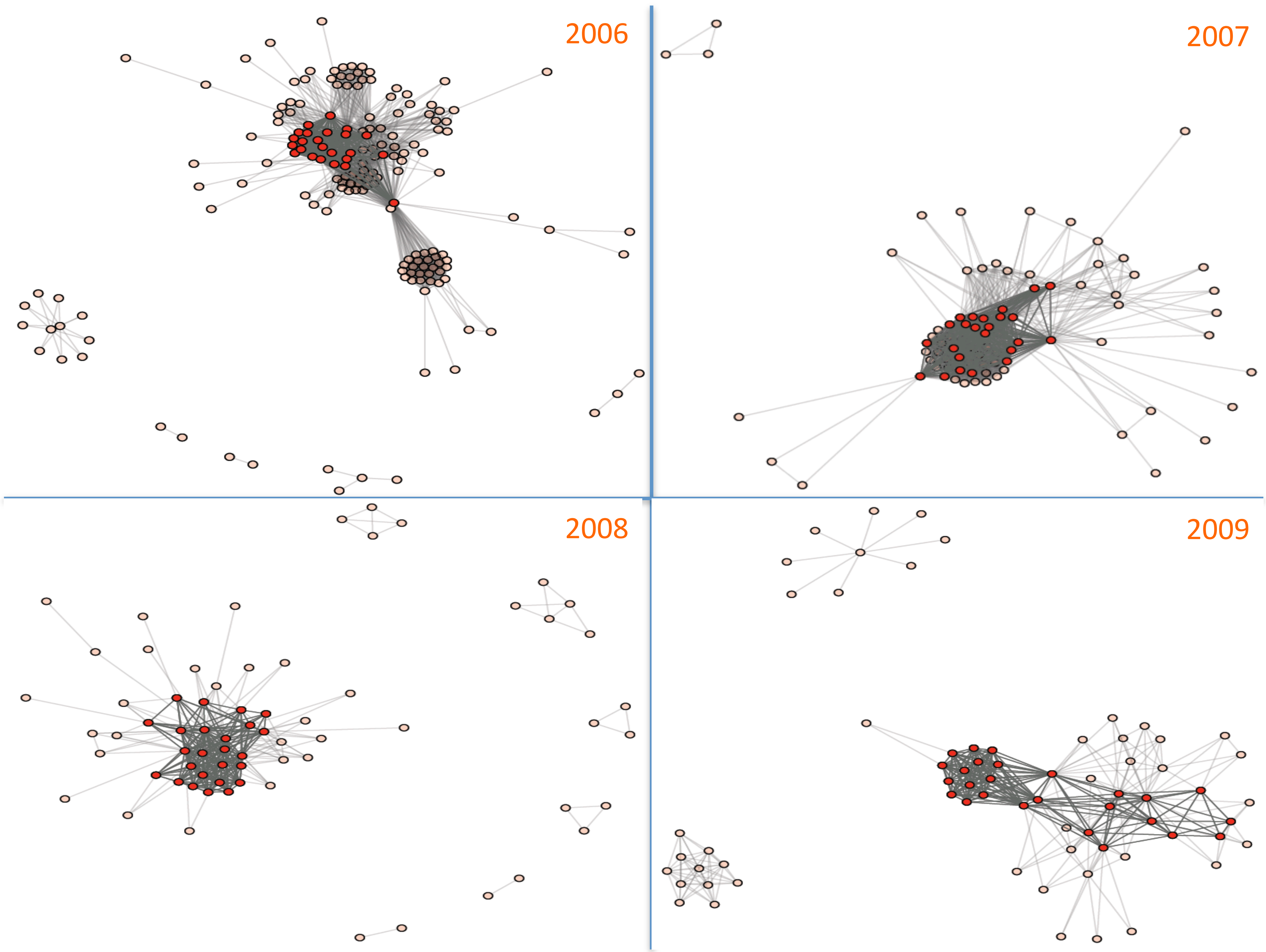
colending.R

```r

#KEEP ONLY THE ACTIVE BANKS
idx = which(colSums(CLmat) > 0)
CLmat = CLmat[idx,idx]
banks = banks[idx]


#COMPUTE EIGENVALUE CENTRALITY
library(igraph); library(Cairo)
g = graph.adjacency(CLmat,mode="undirected",diag=FALSE)
eig_cent = evcent(g)$vector
sol = sort(eig_cent,index.return=TRUE,decreasing=TRUE)
topcent = cbind(sol$ix, banks[sol$ix],eig_cent[sol$ix])
print("Eigenvalue Centrality")
print(topcent[1:min(length(topcent[,1]),25),])
cutoff = as.double( min( topcent[1:min(length(topcent[,1]),25),3] ))
idx = which(eig_cent >= cutoff)

#PRINT SUMMARY STATS
print(c("Number of colending banks = ",length(idx)))
print(c("Number of pairwise relationships = ",sum(CLmat)/2))
print(c("Total coloans = ",num_loans))
P = degree.distribution(g)
d = seq(0,length(P)-1)
expanderR = sum(d^2 *P)/sum(d*P)
print(c("Expander R metric of systemic risk =",expanderR))
diam = diameter(g,directed=FALSE)
print(c("Diameter = ",diam))


#GRAPHING MAIN NETWORK
g = simplify(g)
V(g)$name = seq(vcount(g))
l = layout.fruchterman.reingold(g)
#l = layout.kamada.kawai(g)
#l = layout.circle(g)
l = layout.norm(l, -1,1,-1,1)
#pdf(file="coloan_net.pdf")
#CairoX11()
plot(g, layout=l, vertex.size=3, vertex.label=NA, vertex.color="#ff000033", edge.color="#55555533", edge.arrow.size=0.3, rescale=FALSE, xlim=range(l[,1]), ylim=range(l[,2]) )

```

colending.R

```r

#OVERLAY TOP CENTRALITY NODES ON GRAPH
g2 = subgraph(g, V(g)[idx]-1)
l2 = l[ V(g)[idx], ]
plot(g2, layout=l2, vertex.size=3, vertex.label=V(g2)$name, vertex.color="#ff0000", edge.color="#555555", vertex.label.dist=0.5, vertex.label.cex=0.8, vertex.label.font=2, edge.arrow.size=0.3, add=TRUE, rescale=FALSE )
#plot(g2, layout=l2, vertex.size=3, vertex.label=NA, vertex.color="#ff0000", edge.color="#555555", vertex.label.dist=0.5, vertex.label.cex=0.8, vertex.label.font=2, edge.arrow.size=0.3, add=TRUE, rescale=FALSE )

#DEGREE CENTRALITY
#deg_cent = degree(g)
#sol = sort(deg_cent,index.return=TRUE,decreasing=TRUE)
#topcent = cbind(banks[sol$ix],deg_cent[sol$ix])
#print("Degree Centrality")
#print(topcent[1:min(length(topcent[,1]),25),])

#BETWEENNESS CENTRALITY
#bet_cent = betweenness(g,directed=FALSE)
#sol = sort(bet_cent,index.return=TRUE,decreasing=TRUE)
#topcent = cbind(banks[sol$ix],bet_cent[sol$ix])
#print("Betweenness Centrality")
#print(topcent[1:min(length(topcent[,1]),25),])

#CLOSENESS CENTRALITY
#clo_cent = closeness(g)
#sol = sort(clo_cent,index.return=TRUE,decreasing=TRUE)
#topcent = cbind(banks[sol$ix],clo_cent[sol$ix])
#print("Closeness Centrality")
#print(topcent[1:min(length(topcent[,1]),25),])

```


2005
Citigroup Inc.
J.P. Morgan Chase
Bank of America Corp.
180
47
143
29
213


2006
2007
2008
2009

Network Fragility

- Definition: how quickly will the failure of any one bank trigger failures across the network?
- Metric: expected degree of neighboring nodes averaged across all nodes.

$$R \geq 2$$

$$E(d^2)/E(d) \equiv R, \text{ where } d \text{ stands for the degree of a node.}$$

- Neighborhoods are expected to "expand" when
- Metric: diameter of the network.

Top 25 banks by systemic risk

Year	#Colending banks	#Coloans	Colending pairs	$R = E(d^2)/E(d)$	Diam.
2005	241	75	10997	137.91	5
2006	171	95	4420	172.45	5
2007	85	49	1793	73.62	4
2008	69	84	681	68.14	4
2009	69	42	598	35.35	4

	(Year = 2005)	
Node #	Financial Institution	Normalized Centrality
143	J P Morgan Chase & Co.	1.000
29	Bank of America Corp.	0.926
47	Citigroup Inc.	0.639
85	Deutsche Bank Ag New York Branch	0.636
225	Wachovia Bank NA	0.617
235	The Bank of New York	0.573
134	Hsbc Bank USA	0.530
39	Barclays Bank Plc	0.530
152	Keycorp	0.524
241	The Royal Bank of Scotland Plc	0.523
6	Abn Amro Bank N.V.	0.448
173	Merrill Lynch Bank USA	0.374
198	PNC Financial Services Group Inc	0.372
180	Morgan Stanley	0.362
42	Bnp Paribas	0.337
205	Royal Bank of Canada	0.289
236	The Bank of Nova Scotia	0.289
218	U.S. Bank NA	0.284
50	Calyon New York Branch	0.273
158	Lehman Brothers Bank Fsb	0.270
213	Sumitomo Mitsui Banking	0.236
214	Suntrust Banks Inc	0.232
221	UBS Loan Finance Llc	0.221
211	State Street Corp	0.210
228	Wells Fargo Bank NA	0.198

Topic 3

PORTFOLIO OPTIMIZATION USING R

The research papers for this work are on my web page – just google it.
http://algo.scu.edu/~sanjivdas/research.htm/

1. Das, Markowitz, Scheid, and Statman (JFQA 2010), "Portfolio Optimization with Mental Accounts"
2. Das & Statman (2008), "Beyond Mean-Variance: Portfolio with Structured Products and non-Gaussian returns."

Standard Optimization Problem

Mean

Risk aversion

Covariance matrix

$$\max_{w} \quad w' \mu - \frac{\gamma}{2} w' \Sigma w$$

$$w' \mathbf{1} = 1 \qquad\qquad \mathbf{1} = [1, 1, \ldots, 1]' \in R^n$$

Portfolio weights

SOLUTION:

$$w = \frac{1}{\gamma} \Sigma^{-1} \left[\mu - \left(\frac{\mathbf{1}'\Sigma^{-1}\mu - \gamma}{\mathbf{1}'\Sigma^{-1}\mathbf{1}} \right) \mathbf{1} \right] \in R^n$$

See D, Markowitz, Scheid, Statman (JFQA 2010)

Solution Math

To solve this maximization problem, we set up the Lagrangian with coefficient λ:

$$\text{(A-1)} \qquad \max_{w,\lambda} L = w'\mu - \frac{\gamma}{2} w' \Sigma w + \lambda[1 - w'\mathbf{1}].$$

The first-order conditions are

$$\text{(A-2)} \qquad \frac{\partial L}{\partial w} = \mu - \gamma \Sigma w - \lambda \mathbf{1} = 0,$$

$$\text{(A-3)} \qquad \frac{\partial L}{\partial \lambda} = 1 - w'\mathbf{1} = 0.$$

Note that equation (A-2) is a system of n equations. Rearranging equation (A-2) gives

$$\text{(A-4)} \qquad \Sigma w = \frac{1}{\gamma}[\mu - \lambda \mathbf{1}],$$

and premultiplying both sides of this equation by Σ^{-1} gives

$$\text{(A-5)} \qquad w = \frac{1}{\gamma} \Sigma^{-1} [\mu - \lambda \mathbf{1}].$$

Final solution

The solution here for portfolio weights is not yet complete, as the equation contains λ, which we still need to solve for. Premultiplying equation (A-5) by $\mathbf{1}'$ gives

$$\text{(A-6)} \qquad \mathbf{1}'w = \frac{1}{\gamma}\mathbf{1}'\Sigma^{-1}[\mu - \lambda\mathbf{1}],$$

$$\text{(A-7)} \qquad 1 = \frac{1}{\gamma}[\mathbf{1}'\Sigma^{-1}\mu - \lambda\mathbf{1}'\Sigma^{-1}\mathbf{1}],$$

which can now be solved for λ to get

$$\text{(A-8)} \qquad \lambda = \frac{\mathbf{1}'\Sigma^{-1}\mu - \gamma}{\mathbf{1}'\Sigma^{-1}\mathbf{1}}.$$

Plugging λ back into equation (A-5) gives the closed-form solution for the optimal portfolio weights:

$$w = \frac{1}{\gamma}\Sigma^{-1}\left[\mu - \left(\frac{\mathbf{1}'\Sigma^{-1}\mu - \gamma}{\mathbf{1}'\Sigma^{-1}\mathbf{1}}\right)\mathbf{1}\right] \in R^n.$$

This optimal solution w is an n-vector and is easily implemented, given that it is analytical.

Example: Construction of Portfolios: Available securities

	Expected returns	Standard deviations
Bond	5%	5%
Low-risk stock	10%	20%
High-risk stock	25%	50%

The correlation between the two stocks is 0.2. Other correlations are zero.

Investor goals (sub-portfolios)

Investor Goals & Sub-Portfolios

Goal (Sub-Portfolio)	Current Allocation	Time Horizon	Annualized Return Goal	Total Accumulation Goal
Bequest	$200,000	25 years	26.35%	$69,248,625
Education	$200,000	3 years	12.18%	$282,343
Retirement Account	$600,000	15 years	10.23%	$2,586,118

Sub-portfolios and overall portfolio

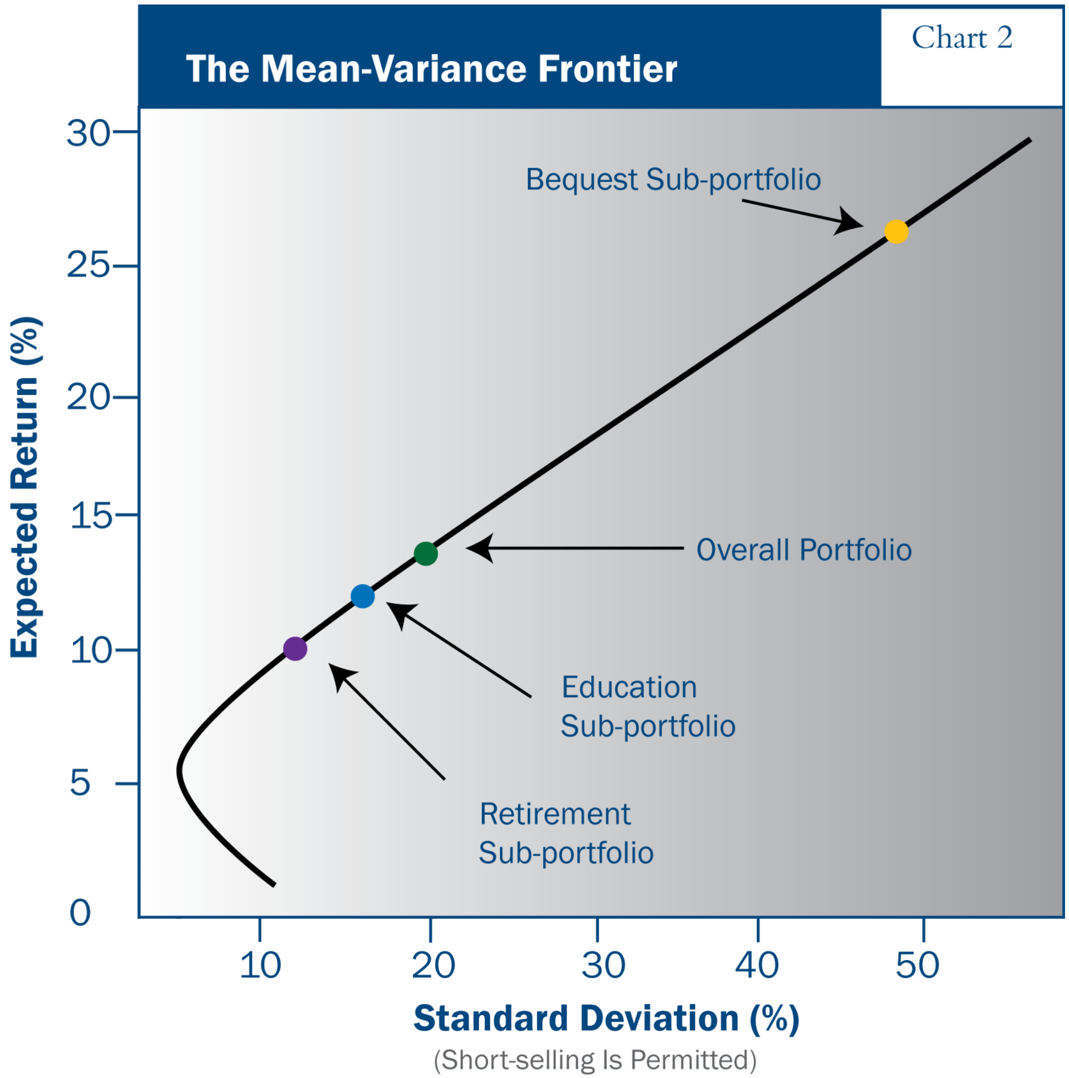
The Three Goal Sub-portfolios and the Overall Portfolio

Assets	Retirement Sub-portfolio	Education Sub-portfolio	Bequest Sub-portfolio	Overall Portfolio
Bond	53.94%	37.87%	(78.90%)	24.16%
Low Risk Stock	26.56%	34.99%	96.20%	42.17%
High Risk Stock	19.50%	27.14%	82.70%	33.67%
Total Weights	100%	100%	100%	100%
Expected Return	10.23%	12.18%	26.35%	13.84%
Std. Deviation	12.30%	16.57%	49.13%	20.32%

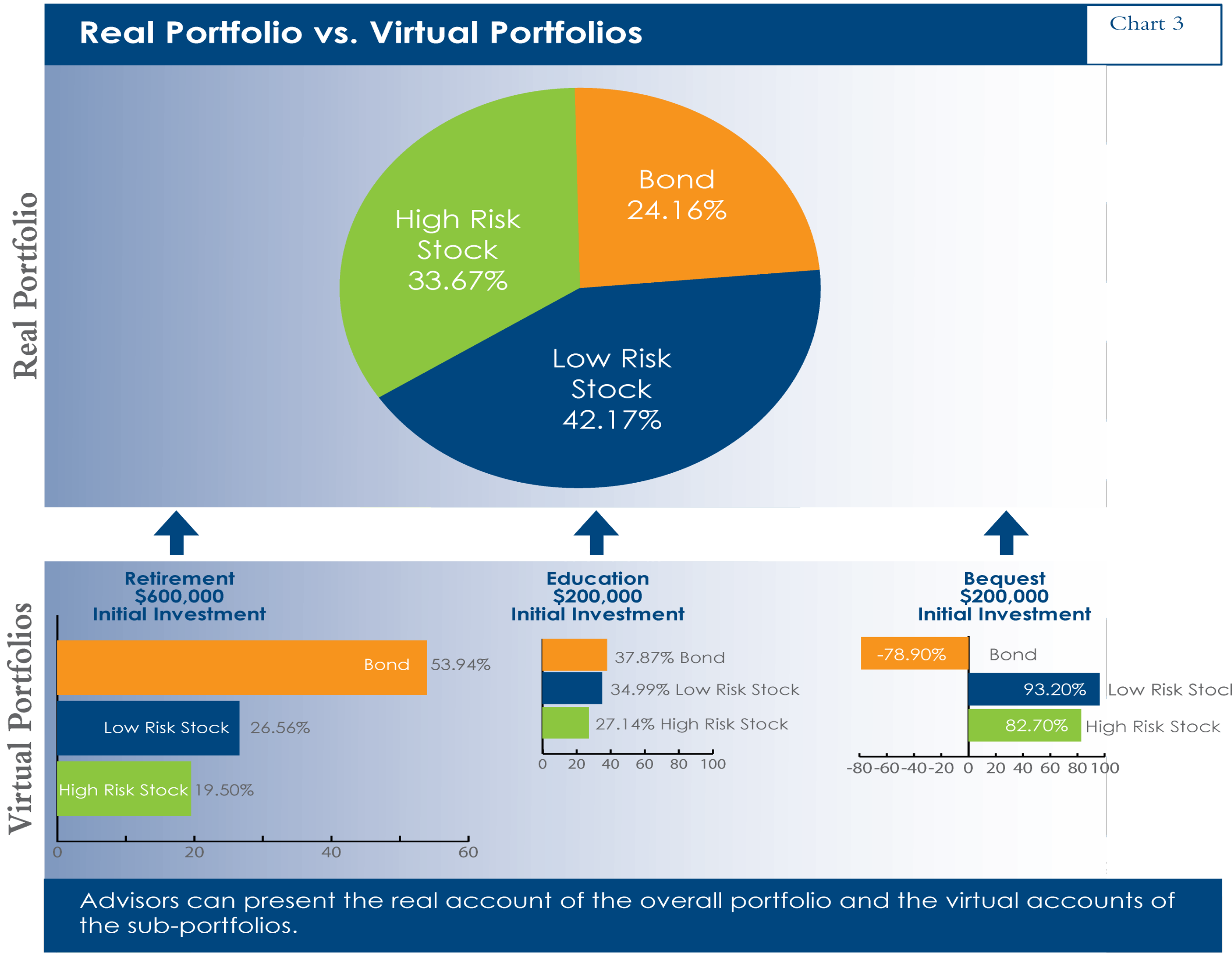
The expected return of the overall portfolio is the weighted average of the expected returns of the sub-portfolios.

The risk of the overall portfolio is *not* the weighted average of the risk of the sub-portfolios.

Mean-variance efficient frontier



Real portfolios versus virtual portfolios



An alternate problem

$$\max_{w} \quad w'\mu, \quad \text{s.t. } \text{Prob}[r \leq H] \leq \alpha$$

$$H \leq w'\mu + \Phi^{-1}(\alpha)[w'\Sigma w]^{1/2}$$

For normal returns

$$w = \frac{1}{\gamma}\Sigma^{-1}\left[\mu - \left(\frac{\mathbf{1}'\Sigma^{-1}\mu - \gamma}{\mathbf{1}'\Sigma^{-1}\mathbf{1}}\right)\mathbf{1}\right] \in R^n$$

Solve for γ

$$H = w(\gamma)'\mu + \Phi^{-1}(\alpha)[w(\gamma)'\Sigma\, w(\gamma)]^{1/2}$$

mv_var.R

```r
#MENTAL ACCOUNT OPTIMIZERS
#Sanjiv Das, July 2007
# This is basically the mean-variance problem with the VaR
# and short selling constraint
# Min beta/2 * Var(p), st E(rp)=ER,
# H < E(rp) + phiinv(alpha)*Std(p)
# w > 0 (for all assets)
# sum(w)=1

library(quadprog)
library(minpack.lm)

#------------------------------------------------------------------
#Function for the basic Markowitz Model
markowitz = function(mu,cv,Er) {
    n = length(mu)
    wuns = matrix(1,n,1)
    A = t(wuns) %*% solve(cv) %*% mu
    B = t(mu) %*% solve(cv) %*% mu
    C = t(wuns) %*% solve(cv) %*% wuns
    D = B*C - A^2
    lam = (C*Er-A)/D
    gam = (B-A*Er)/D
    wts = lam[1]*(solve(cv) %*% mu) + gam[1]*(solve(cv) %*% wuns)
}
```

mv_var.R

```r
#----------------------------------------------------------------
#Function for standard Markowitz model with no short sales allowed
# min  1/2 * (w' cv w)
# st:    w' 1 = 1
#        w' mu = Er
#        w >= 0
markowitz_nss = function(mu,cv,Er) {
    n = length(mu)
    wuns = matrix(1,n,1)    #For the fully invested constraint
    zers = matrix(0,n,1)    #For the short sale lower bound on wts
    Bmat = diag(n)          #Setting up the w>LB constraints
    Amat = cbind(mu,wuns,Bmat)  #Complete lhs constraint matrix
    bvec = c(Er,1,t(zers))      #rhs of constraints
    dvec = matrix(0,n,1)
    sol = solve.QP(cv,dvec,Amat,bvec,meq=2)
}

#----------------------------------------------------------------
#Function for constrained Markowitz model in beta form
# max w'*mu - beta/2 * (w' cv w)
# st:    w' 1 = 1
#        w >= LB
#        w <= UB
mv_constr = function(mu,cv,LB,UB,beta) {
    n = length(mu)
    bcv = beta * cv         #Multiply the cv by beta for obj fn
```

Risk as probability of losses

Sub-portfolio Risk Defined by the Probability of Losses

	Retirement Sub-portfolio	Education Sub-portfolio	Bequest Sub-portfolio
Expected Return	10.23%	12.18%	26.35%
Std. Deviation	12.30%	16.57%	49.13%
Probability-Based Risk Language	No more than a 5% probability of losing more than 10%	No more than a 15% probability of losing more than 5%	No more than a 20% probability of losing more than 15%

Mean-variance problem: Minimize Risk (variance) subject to minimum level of Expected Return.

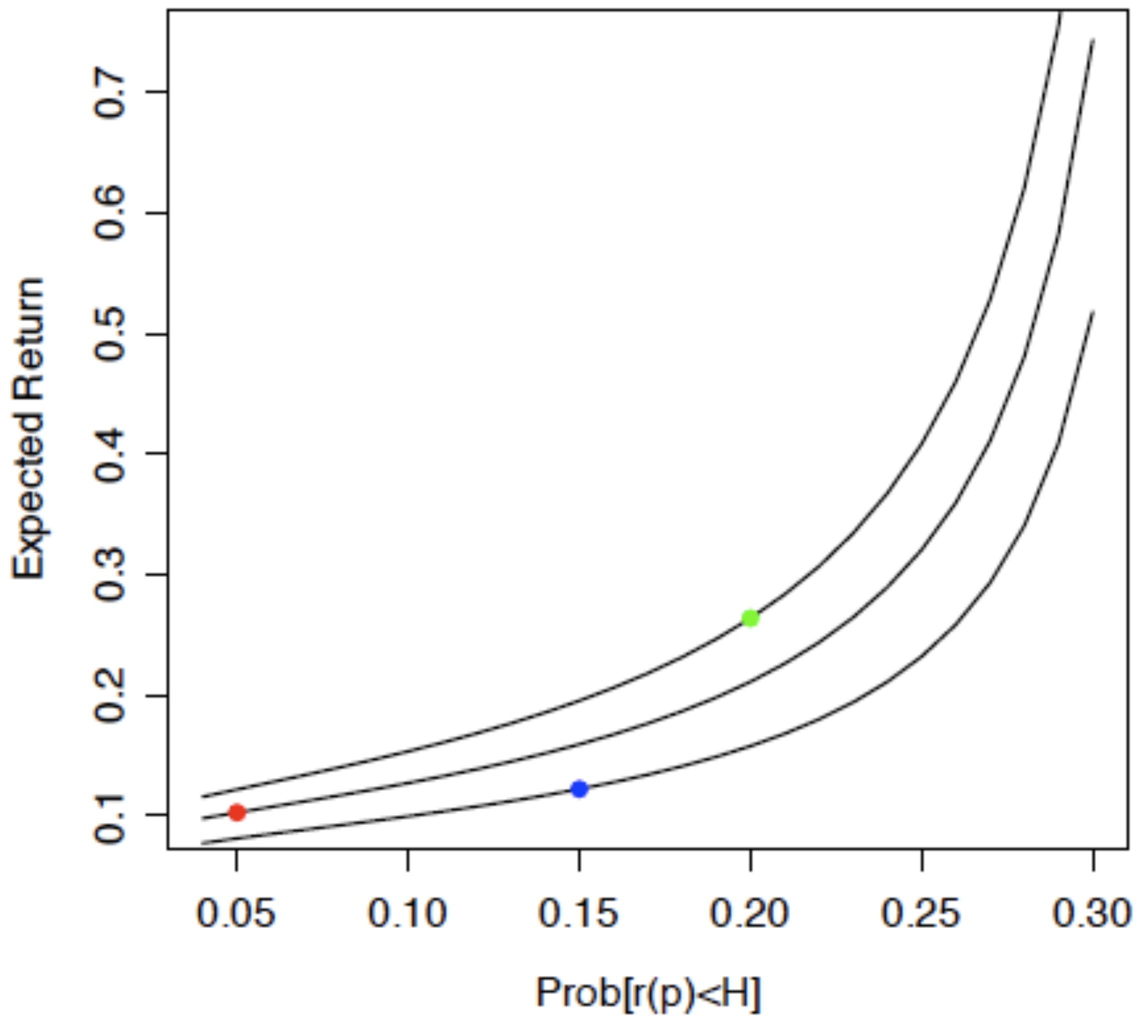
↕

Behavioral portfolio theory: Maximize Return subject to a maximum probability of falling below a threshold.

Subportfolio Scenarios

Risk aversion:	$\gamma = 3.7950$	$\gamma = 2.7063$	$\gamma = 0.8773$	60:20:20 mix
	Retirement	Education	Bequest	Aggregate
	Sub-portfolio	Sub-portfolio	Sub-portfolio	Portfolio
Threshold (H)	Prob[$r < H$]	Prob[$r < H$]	Prob[$r < H$]	Prob[$r < H$]
-25.00%	0.00	0.01	0.15	0.03
-20.00%	0.01	0.03	0.17	0.05
-15.00%	0.02	0.05	0.20	0.08
-10.00%	0.05	0.09	0.23	0.12
-5.00%	0.11	0.15	0.26	0.18
0.00%	0.20	0.23	0.30	0.25
5.00%	0.34	0.33	0.33	0.33
10.00%	0.49	0.45	0.37	0.42
15.00%	0.65	0.57	0.41	0.52
20.00%	0.79	0.68	0.45	0.62
25.00%	0.89	0.78	0.49	0.71
Mean return	10.23%	12.18%	26.35%	13.84%
Std. deviation	12.30%	16.57%	49.13%	20.32%

Efficient Frontiers in the BPT (Mental Account) World



Short-Selling Constraints

$$
\begin{array}{rcl}
\max\limits_{w} \quad w'\mu & & s.t. \\
w'\mathbf{1} & = & 1 \\
w & \geq & L \\
w & \leq & U \\
w'\mu + \Phi^{-1}(\alpha)\sqrt{w'\Sigma w} & \geq & H
\end{array}
$$

Linear program with *non-linear* constraints. This is not a standard quadratic programming problem (QP) like the Markowitz model.

MVT uses a standard QP: *quadratic* objective function with *linear* constraints.

Modified Problem

$$\text{Solve}_\gamma \quad w(\gamma)'\mu + \Phi^{-1}(\alpha)\sqrt{w(\gamma)'\Sigma w(\gamma)} \quad = \quad H,$$

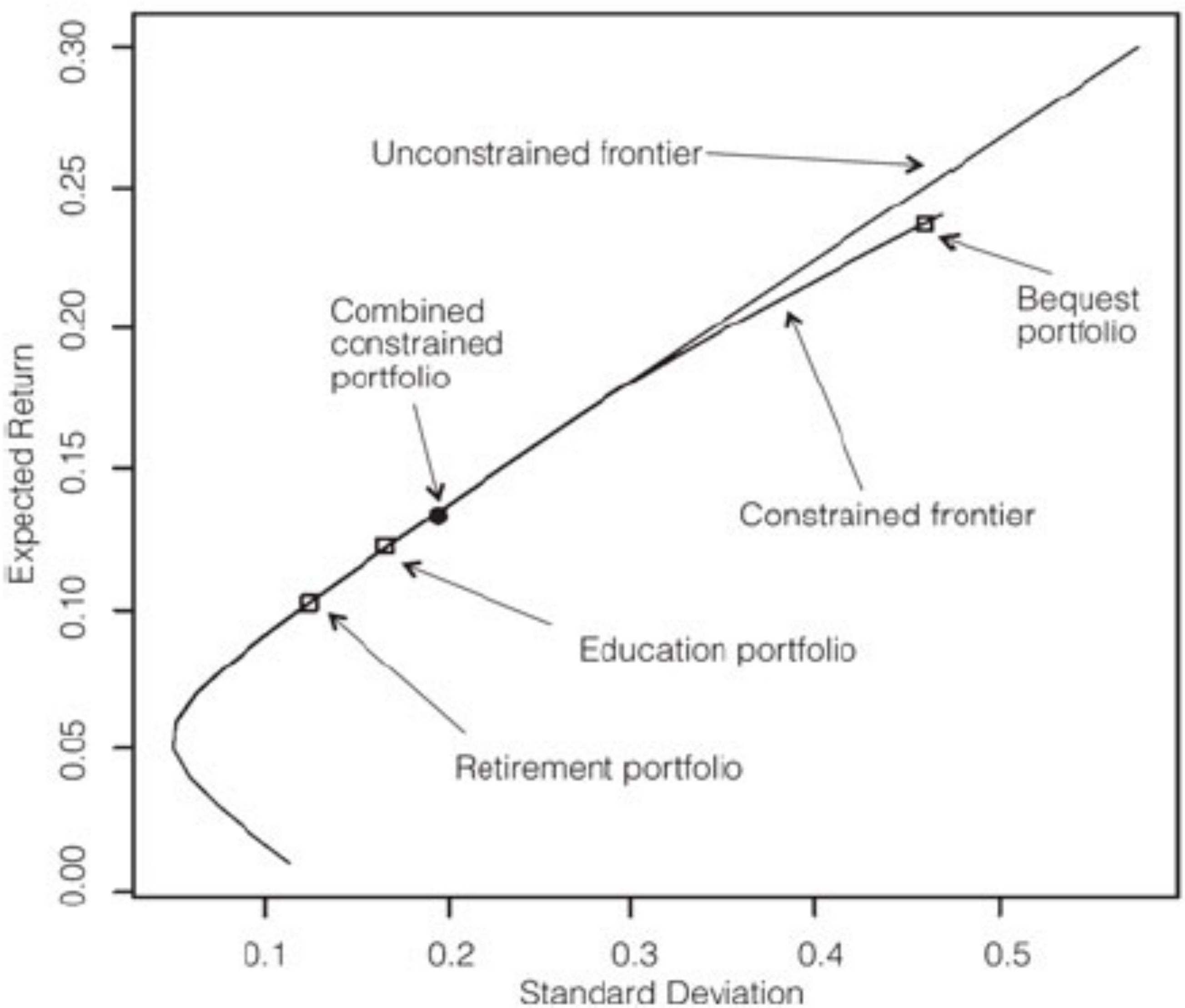
$w(\gamma)$ is the solution to the following optimization program:

$$\begin{aligned} \max_{w} \quad w'\mu - \frac{\gamma}{2}w'\Sigma w \qquad & \qquad \text{s.t.} \\ w'\mathbf{1} \quad = & \quad 1, \\ w \quad \geq & \quad L, \\ w \quad \leq & \quad U. \end{aligned}$$

Standard QP problem
with linear constraints

Amenable to industrial optimizers; we use the R system with the **quadprog** package and **minpack.lm** library.

MV Frontier with Short-selling



Deviating from Normality with Copulas

$$
\begin{aligned}
F(r_1, r_2, \ldots, r_n) &= C[F_1(r_1), F_2(r_2), \ldots, F_n(r_n)] \\
&= C(u_1, u_2, \ldots, u_n) \\
f(r_1, \ldots, r_n) &= \frac{\partial^n F}{\partial r_1 \ldots \partial r_n} \\
&= \frac{\partial C}{\partial u_1 \ldots \partial u_n} \times \frac{\partial u_1}{\partial r_1} \times \ldots \times \frac{\partial u_n}{\partial r_n} \\
&= c(u_1, \ldots, u_n) \times f_1(r_1) \times \ldots \times f_n(r_n) \\
&= c(u_1, \ldots, u_n) \prod_{i=1}^{n} f_i(r_i)
\end{aligned}
$$

Gaussian Copula

$$C_\rho(u_1,\ldots,u_n) = \Phi[\Phi^{-1}(u_1),\ldots,\Phi^{-1}(u_n)]$$

$$\begin{aligned}
c_\rho(u_1,\ldots,u_n) &= \frac{\partial^n C_\rho[F_1(r_1),F_2(r_2),\ldots,F_n(r_n)]}{\partial r_1\ldots\partial r_n} \\
&= \frac{f(r_1,\ldots,r_n)}{\prod_{i=1}^n f_i(r_i)} \\
&= \frac{1}{\sqrt{|\rho|}}\exp\left[-\frac{1}{2}\mathbf{r}'(\rho^{-1}-\mathbf{I})\mathbf{r}\right]
\end{aligned}$$

$$f(\mathbf{r}) = \frac{1}{\sqrt{|\rho|}}\exp\left[-\frac{1}{2}\mathbf{r}'(\rho^{-1}-\mathbf{I})\mathbf{r}\right] \times \prod_{i=1}^n f_i(r_i)$$

Extended Optimization Problem

$$
\begin{aligned}
\max_{w \in R^n} \quad E[w'\mathbf{r}] \equiv & \quad \int_{\mathbf{r}} [w' \; \mathbf{r}] \; \cdot \; f(\mathbf{r}) \; \cdot \; d\mathbf{r}, \\
\text{subject to} & \\
w'\mathbf{1} &= 1 \\
w &\geq L_l \\
w &\leq L_u \\
\text{Prob}[w'\mathbf{r} \leq H] &\leq \alpha
\end{aligned}
$$

$$
f(\mathbf{r}) = \frac{1}{\sqrt{|\rho|}} \exp\left[-\frac{1}{2}\mathbf{r}'(\rho^{-1} - \mathbf{I})\mathbf{r}\right] \times \prod_{i=1}^{n} f_i(r_i)
$$

```r
#COPULA MV PROBABILITY FUNCTION
#USES CREATE_COPULA_DENSITY and CREATE_COPULA_MARGINALS
#Works for up to 2-4 assets (easily extended to more assets w/ more code)
#cv is nxn, df relate to T-distrbn
#m is the number of values on the support of each asset
copula_prob = function(cv,df,m,zlim) {
    n = length(cv[,1])
    z = seq(-zlim,zlim,len=m)
    #Expand grid depending on how many assets there are
    if (n==2) { r = as.matrix(expand.grid(z,z)) }
    if (n==3) { r = as.matrix(expand.grid(z,z,z)) }
    if (n==4) { r = as.matrix(expand.grid(z,z,z,z)) }
    #Compute correlation matrix
    sig = sqrt(diag(cv))
    sigk = kronecker(t(sig),sig)
    cr = cv/sigk    #Correlation matrix = cr
    #Call copula density and marginals
    copdens = create_copula_density(cr,n,m,r)
    copmarg = create_copula_marginals(df,n,m,r)
    dz = z[2]-z[1]
    probs = (copdens * copmarg) * (dz^n)
    probs = probs/sum(probs)
}
```

```r
#CALL OPTIMIZER
w = results[idxmax,1:3]     #Starting parameter values
w = 0.999*w/sum(w)+0.000001 #Make sure the start values satisfy constraint
#sol = optim(w,objfn,method="L-BFGS-B",lower=c(0,0,0),upper=c(1,1,1),ret=R,pr=pr,H=H,alpha=alpha)
#sol = optim(w,objfn,ret=R,pr=pr,H=H,alpha=alpha)  #if there are no bounds
sol = constrOptim(w,objfn,grad=NULL,ui=rbind(c(-1,-1,-1),c(1,0,0),c(0,1,0),c(0,0,1)),ci=c
(-1,0,0,0),outer.eps=1e-09,ret=R,pr=pr,H=H,alpha=alpha)
print(sol)
w = sol$par
#w[length(w)] = 1-sum(w[1:(length(w)-1)])
print("Optimal weights"); print(w)
momts=moments4return(w,R,pr); print("Optimized moments"); print(momts)
probh=prob_fail(w,R,pr,H); print("Prob[r<H]"); print(probh)
result_matrix = rbind(result_matrix,c(0,w,0,probh,momts,0,0,0))
```

Solution

Table 1: Optimal portfolio allocations when the distribution of joint returns is Student-t. Using the same input data as in section 2.3, the portfolio is optimized when the asset distributions are multivariate Student-t, with declining degrees of freedom (dof in the table) (i.e. increasing tail fatness). When $dof = \infty$, the joint distribution converges to the limiting multivariate normal. The threshold return is $H = -0.10$ and the permissible probability of failing to reach this threshold is $\alpha = 0.05$. The portfolio expected return is maximized subject to the constraint that the the maximum probability of failing to reach the threshold H is α. We report the weights in each of the three assets: $\{w_1, w_2, w_3\}$ and the expected return of the optimized portfolio. The weight of the low risk asset is w_1, that of the medium risk asset is w_2 and of the high risk asset is w_3.

	Risk of Asset			
	Low	Medium	High	
dof	w_1	w_2	w_3	$E(r)\%$
1000	0.58	0.21	0.21	10.20
50	0.59	0.21	0.20	10.02
10	0.60	0.25	0.15	9.30
5	0.79	0.06	0.15	8.24

Non-Linear Products

$$\max_{w} \int_{u \in U} [w' \, r(u)] \; \cdot \; p[r(u)] \; \cdot \; \Pi_{i=1}^{n} dr_i(u)$$

U is a set of states over
n assets

r(u) is a n-vector of random returns

$$\sum_{i=1}^{n} w_i \;\; = \;\; 1, \quad w_i \geq 0$$

$$\int_{u:[w' r(u)]] < H} p[r(u)] \; \cdot \; \Pi_{i=1}^{n} dr_i(u) \;\; \leq \;\; \alpha$$

Assume that of the n securities, n' are "primary" assets
$(n - n')$ securities are "derivative" assets

Formally, $U' = \{r(u)\}, u = 1, \ldots, m^{n'}$. It is then easy to expand this state space U' into the full state space U by augmenting the space with the returns of the derivative assets, that are easily written as functions of the primary assets. We note that U may be represented by a matrix of returns, i.e., $U \in R^{m^{n'} \times n}$. The returns of all assets in each state are given by the rows of this matrix.

Compute p[r(u)]

Restatement of the problem

$$\max_{w} \int_{u \in U} [w' \, r(u)] \cdot p[r(u)] \cdot \Pi_{i=1}^{n} dr_i(u)$$

$$-X \left[\alpha - \int_{u:[w' r(u)]] < H} p[r(u)] \; \Pi_{i=1}^{n} dr_i(u) \right]^2$$

$$\text{where} \quad w_n = 1 - \sum_{i=1}^{n-1} w_i$$

This is a quadratic optimization with linear constraints.

Not a quadratic optimization with non-linear constraints.

Introducing Structured Products

Can we improve the risk-adjusted returns in a portfolio by using puts and calls?

Derivatives are very risky.

Optimal mental account allocations when put options are included as the fourth security. The top panel of the table shows basic statistics of the returns of puts. The threshold return is varied in the range $H = \{-10\%, -5\%, 0, +1\%, +2\%\}$ and the maximum probability of failing to reach these thresholds is $\alpha = 0.05$. The risk free rate is set to 3%. The expected return of the mental account is maximized subject to the constraint that the maximum probability of failing to reach the threshold H is α. We report the weights in each of the four securities: $\{w_1, w_2, w_3, w_4\}$ (corresponding to the low, medium, and high-risk securities, and the put), and the expected return and standard deviation of the optimized mental account. The first line of the table shows the mental account when puts are not available and the remaining lines show it when they are.

Panel *A*: Put Return Statistics

Strike	Put Price	Put Mean	Put Variance
0.8	0.0086	-0.3114	11.8849
0.9	0.0277	-0.3941	3.5614
1	0.0646	-0.3847	1.6320
1.1	0.1204	-0.3358	0.9379
1.2	0.1922	-0.2730	0.5980
1.3	0.2752	-0.2122	0.3962
1.4	0.3649	-0.1617	0.2665

And so ….

Are puts optimal?

Panel B: $H = -10\%, \alpha = 0.05$

	LowRisk	MedRisk	HighRisk	LongPut		Portfolio Return Moments			
Strike	w_1	w_2	w_3	w_4	$Pr[r < H]$	Mean	Std Dev	Skewness	Kurtosis
No put	0.5871	0.2052	0.2077	-	0.0498	0.1018	0.1226	0.0000	0.0000
0.8	0.5773	0.2058	0.2139	0.0029	0.0497	0.1020	0.1235	0.0335	-0.0184
0.9	0.5862	0.2054	0.2082	0.0003	0.0496	0.1018	0.1227	0.0031	-0.0020
1	0.5873	0.2034	0.2078	0.0000	0.0496	0.1017	0.1225	0.0001	0.0000
1.1	0.5796	0.2112	0.2081	0.0011	0.0498	0.1018	0.1229	0.0046	0.0001
1.2	0.5897	0.2011	0.2091	0.0000	0.0499	0.1019	0.1229	0.0000	0.0000

But …

Puts are needed when the threshold return is high

Panel C: $H = -5\%, \alpha = 0.05$

	LowRisk	MedRisk	HighRisk	LongPut		Portfolio Return Moments			
Strike	w_1	w_2	w_3	w_4	$Pr[r < H]$	Mean	Std Dev	Skewness	Kurtosis
No put	0.6964	0.2008	0.1028	0.0000	0.0498	0.0806	0.0793	0.0000	0.0000
0.8	0.6986	0.1998	0.1016	0.0000	0.0491	0.0803	0.0788	0.0008	-0.0010
0.9	0.6982	0.2002	0.1016	0.0000	0.0491	0.0803	0.0788	0.0003	-0.0002
1	0.6971	0.2000	0.1028	0.0001	0.0495	0.0805	0.0792	0.0011	-0.0005
1.1	0.6962	0.2000	0.1035	0.0003	0.0496	0.0806	0.0793	0.0032	0.0000
1.2	0.6977	0.1994	0.1029	0.0000	0.0495	0.0805	0.0792	0.0000	0.0000

Panel D: $H = 0\%, \alpha = 0.05$

	LowRisk	MedRisk	HighRisk	LongPut		Portfolio Return Moments			
Strike	w_1	w_2	w_3	w_4	$Pr[r < H]$	Mean	Std Dev	Skewness	Kurtosis
No put	Infeasible: No solution								
0.8-1.1	Infeasible: No solution								
1.2	0.0488	0.7996	0.0043	0.1473	0.0495	0.0433	0.0688	2.4538	6.5319
1.3	0.2459	0.6007	0.0104	0.1430	0.0495	0.0446	0.0422	2.4319	8.7183

For high thresholds the investor cannot get an acceptable portfolio without puts.

Should investors use calls?

Calls are risky too.

The expected return of the call ranges from 32% to 51%.

The variance of return is very large. For example, the one-standard deviation of return for the at-the-money option is 158%($= \sqrt{249\%}$, square-root of the variance).

Panel *A*: Call Return Statistics			
Strike	Call Price	Call Mean	Call Variance
0.8	0.2322	0.3173	0.6582
0.9	0.1543	0.4050	1.2599
1	0.0941	0.4844	2.4928
1.1	0.0529	0.5111	4.8543
1.2	0.0277	0.4362	8.8926

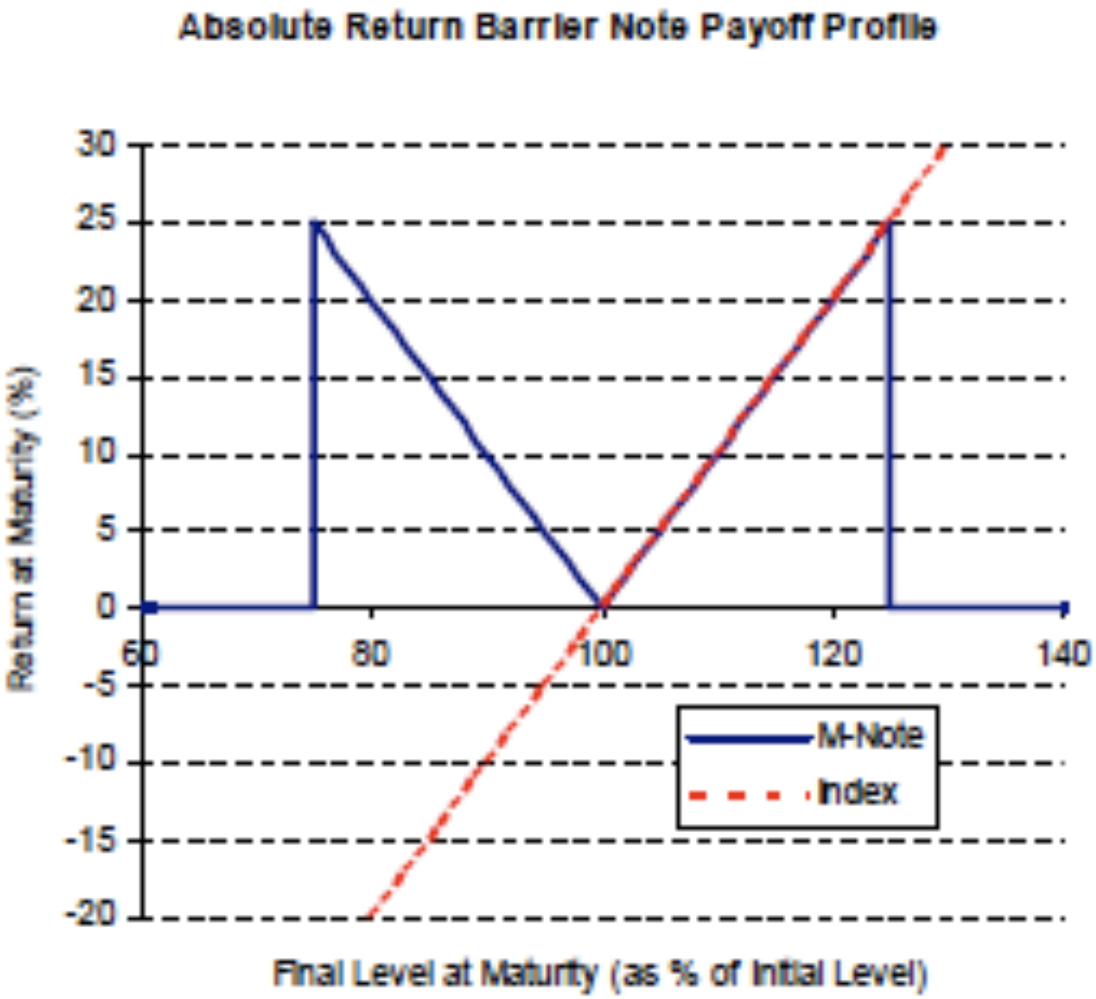
But have attractive and high mean returns!

Calls give better portfolios

Panel B: $H = -10\%, \alpha = 5\%$									
	LowRisk	MedRisk	HighRisk	LongCall		Portfolio Return Moments			
Strike	w_1	w_2	w_3	w_4	$Pr[r < H]$	Mean	Std Dev	Skewness	Kurtosis
No Call	0.5871	0.2052	0.2077	0.0000	0.0498	0.1018	0.1226	0.0000	0.0000
0.8	0.7419	0.0027	0.1524	0.1030	0.0496	0.1081	0.1296	0.1186	-0.1196
0.9	0.7968	0.0008	0.1005	0.1019	0.0496	0.1063	0.1395	0.3934	-0.1643
1	0.6978	0.1000	0.1498	0.0514	0.0496	0.1072	0.1403	0.4424	0.0592
1.1	0.7954	0.0032	0.1483	0.0531	0.0497	0.1043	0.1546	0.9297	1.0514
1.2	0.5948	0.1995	0.2017	0.0040	0.0495	0.1019	0.1244	0.0817	0.0479

Improvement is greater than 60 bps !

Structured Product: The Barrier-M-note



Barrier-M Note

$$r_4 = \begin{cases} |r_2| & \text{if } |r_2| \leq 0.25 \\ 0 & \text{if } |r_2| > 0.25 \end{cases}$$

(1) A long call at strike \$1.
(2) A long put at strike \$1.
(3) A short call at strike $\$(1+M)$.
(4) A short put at strike $\$(1-M)$.
(5) M short cash-or-nothing calls at strike $\$(1+M)$.
(6) M short cash-or-nothing puts at strike $\$(1-M)$.

$$CON_{call}[\text{Strike} = 1+M] = e^{-rT} N\left[\frac{\ln(1/(1+M)) + (r-\sigma^2/2)T}{\sigma\sqrt{T}}\right]$$

$$CON_{put}[\text{Strike} = 1-M] = e^{-rT} N\left[-\left(\frac{\ln(1/(1-M)) + (r-\sigma^2/2)T}{\sigma\sqrt{T}}\right)\right]$$

"Truncated Straddle"

Barrier-M-note

Table 7

Optimal portfolio allocations when Barrier-M notes are added to the initial three securities. The expected return of a mental account is maximized subject to the constraint that the maximum probability of failing to reach the threshold $H = -10\%$ is $\alpha = 0.05$. We report the weights in each of the four securities: $\{w_1, w_2, w_3, w_4\}$ (corresponding to Low, Medium, and high-risk securities, and the barrier note), and the expected return and standard deviation of the optimized mental account. The first line of the table shows the mental account when Barrier-M notes are not allowed and the second line when they are. The expected return of the note is 7.51% and its standard deviation is 7.06%.

	$H = -10\%, \alpha = 0.05$								
	LowRisk	MedRisk	HighRisk	Long Note		Portfolio Return Moments			
Strike	w_1	w_2	w_3	w_4	$Pr[r < H]$	Mean	Std Dev	Skewness	Kurtosis
No note	0.5871	0.2052	0.2077	-	0.0498	0.1018	0.1226	0.0000	0.0000
Barrier: ±0.25	0.0009	0.0426	0.2635	0.6929	0.0495	0.1237	0.1405	0.0131	0.0048

Return pick-up greater than 250 bps!

Equity-Indexed Product

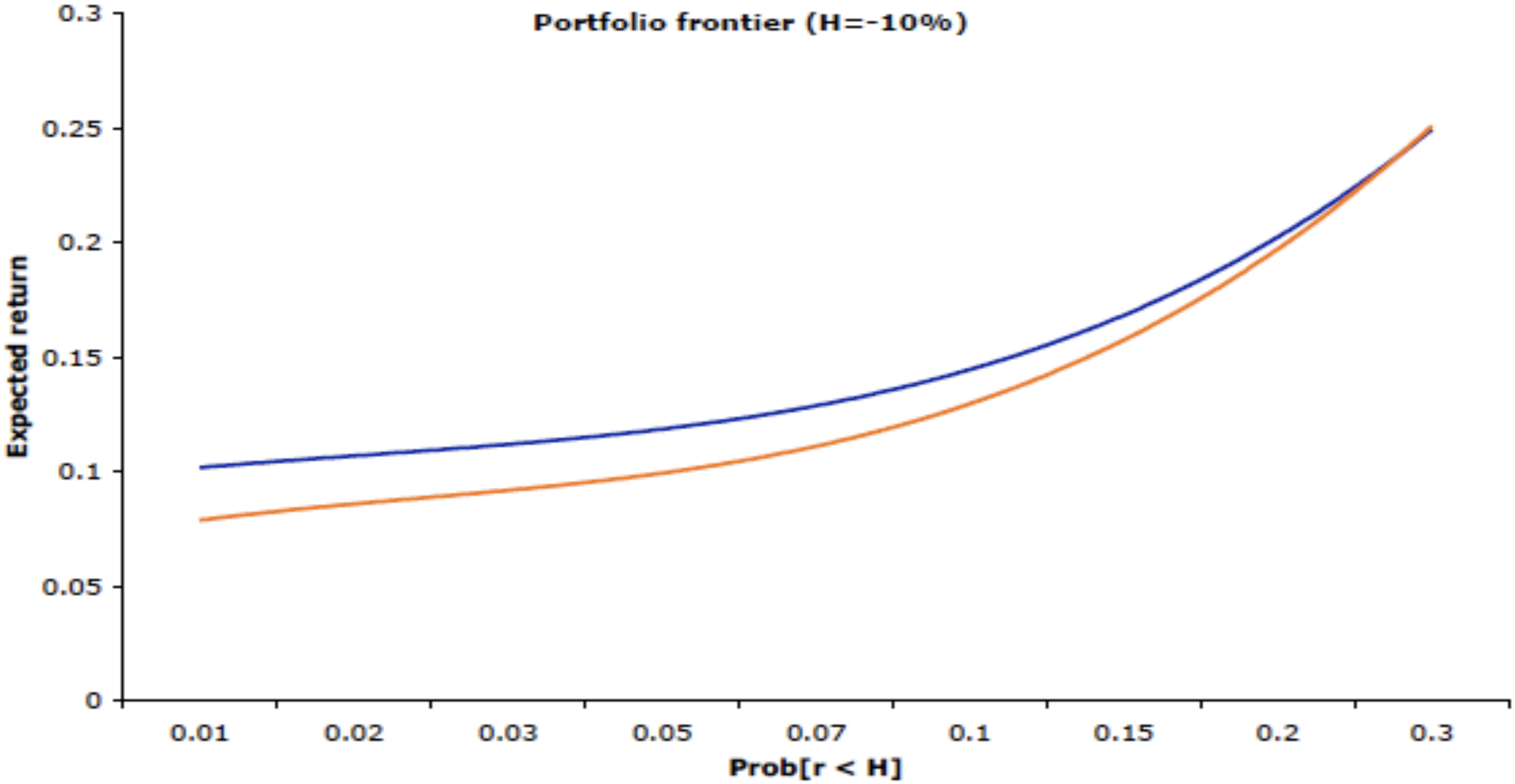


Figure 4: Risk-return frontiers with and without annuities in the portfolio. The graph plots represent behavioral portfolio frontiers. The plot has risk on the x-axis, represented by the probability of portfolio returns dropping below a given threshold of -10%. Expected returns of the portfolio are shown on the y-axis. The specific annuity used is one that has capital protection and an upside cap of 15% with no fees. The higher line is for the portfolio with annuities and the lower line for the portfolio without annuities. As risk increases expected returns increase as well.

Conclusion

- Investors find it easier to think in terms of mental accounts or sub-portfolios when trying to reach their separate financial goals.
- Behavioral portfolio theory deals with maximizing return subject to managing the risk of loss. This problem has a mathematical mapping into mean-variance optimization, yet is much more general.
- Even with short-selling prohibited, the loss from sub-portfolio optimization is smaller than the loss from misestimating investor preferences.
- Reporting performance by sub-portfolio enables investors to track their goals better.
- Goal-based optimization enables choosing portfolios even when normality is not assumed.
- Goal-based optimization provides a framework for including structured products in investor portfolios.

The research papers for this work are on my web page – just google it.
http://algo.scu.edu/~sanjivdas/research.htm/

1. Das, Markowitz, Scheid, and Statman (JFQA 2010), "Portfolio Optimization with Mental Accounts"
2. Das & Statman (2008), "Beyond Mean-Variance: Portfolio with Structured Products and non-Gaussian returns."

Topic 4

PEDAGOGICAL USES FOR R USING THE WEB

http://sanjivdas.wordpress.com/

Use the Rcgi package from David Firth: http://www.omegahat.org/CGIwithR/

You need two program files to get everything working.

(a) The html file that is the web form for input data.

(b) The R file, with special tags for use with the CGIwithR package.

But first, let's create the html file for the web page that will take these three input values. We call it "mortgage_calc.html". The code is all standard, for those familiar with html, and even if you are not used to html, the code is self-explanatory.

```
01 <html>
02 <head>
03 <title>Monthly Mortgage Payment Calculator</title>
04 </head>
05
06 <FORM action="/cgi-bin/R.cgi/mortgage_calc.R" method="POST">
07 <body>
08 Loan Principal: <INPUT name="L" value="" size=5><p>
09 Annual Loan Rate: <INPUT name="rL" value="" size=5><p>
10 Remaining months: <INPUT name="N" value="" size=5><p>
11
12 <P><INPUT type="submit" size=3>
13
14 </body>
15 </html>
```

Notice that line 06 will be the one referencing the R program that does the calculation. The three inputs are accepted in lines 08–10. Line 12 sends the inputs to the R program.

R Code called from CGI

```
#! /usr/bin/R

tag(HTML)
    tag(HEAD)
        tag(TITLE)
            cat("Mortgage Monthly Payment Calculator")
        untag(TITLE)
    untag(HEAD)

tag(h3)
    cat("Mortgage Monthly Payment Calculator")
untag(h3)

lf(2)
tag(BODY)

tag(p)
    tag(b)
        cat("Inputs:")
    untag(b)

    tag(p)
    L = as.numeric(scanText(formData$L))
    cat("Loan Principal: ")
    cat(L)

    tag(p)
    rL = as.numeric(scanText(formData$rL))
    cat("Annual Loan Rate: ")
    cat(rL)

    tag(p)
    N = as.numeric(scanText(formData$N))
    cat("Remaining months: ")
    cat(N)
untag(p)

lf(2)
tag(p)
    cat("Monthly Loan Payment: ")
untag(p)

r = rL/12
mp = r*L/(1-(1+r)^(-N))
cat(mp)

untag(BODY)
untag(HTML)
```

1. Make sure that your Mac is allowing connections to its web server. Go to System Preferences and choose Sharing. In this window enable Web Sharing by ticking the box next to it.
2. Place the html file "mortgage_calc.html" in the directory that serves up web pages. On a Mac, there is already a web directory for this called "Sites". It's a good idea to open a separate subdirectory called (say) "Rcgi" below this one for the R related programs and put the html file there.
3. The R program "mortgage_calc.R" must go in the directory that has been assigned for CGI executables. On a Mac, the default for this directory is "/Library/WebServer/CGI-Executables" and is usually referenced by the alias "cgi-bin" (stands for cgi binaries). Drop the R program into this directory.
4. Two more important files are created when you install the "Rcgi" package. The CGIwithR installation creates two files:
   (a) A hidden file called ".Rprofile"
   (b) A file called R.cgi

   Place both these files in the directory: /Library/WebServer/CGI-Executables

   If you cannot find the .Rprofile file then create it directly by opening a text editor and adding two lines to the file:

   ```
   #! /usr/bin/R
   library(CGIwithR,warn.conflicts=FALSE)
   ```

   Now, open the R.cgi file and make sure that the line pointing to the R executable in the file is showing

   ```
   R_DEFAULT=/usr/bin/R
   ```

   The file may actually have it as "#! /usr/local/bin/R" which is for Linux platforms, but the usual Mac install has the executable in "#! /usr/bin/R" so make sure this is done.

   Make both files executable as follows:

   ```
   chmod a+rx .Rprofile
   chmod a+rx R.cgi
   ```
5. Finally, make the ~/Sites/Rcgi/ directory write accessible:

   ```
   chmod a+wx ~/Sites/Rcgi
   ```

Just being patient and following all the steps makes sure it all works well. Having done it once, it's easy to repeat and create several functions. You can try this example out on my web server at the following link.

http://algo.scu.edu/~sanjivdas/Rcgi/mortgage_calc.html

High-performance computing (parallelR)

Network modeling

Optimization

Q?

Web functions

Calling C from R

Lattice dynamic optimization

High-dimensional distributions with copulas