Section 1

Short Term Variability
Short term market variability—on a time scale from one day to one or two weeks—imposes risk on market participants. One aspect of this risk is that adverse changes in market conditions may result in unacceptable losses, that may lead to the dissolution of a “desk” or to the bankruptcy of an institution. These are the direct impacts of market risk on the institution. A second aspect, less direct but no less dangerous, is that such market risk may lead to the default of one of the institution’s counterparties and consequent loss to the institution. This is called credit risk. We deal first with direct market risk.
Market Risk

Suppose that a bank has a portfolio of financial instruments subject to market risk. These may be actual securities or derivatives. Write $V_i(t, m)$ for the value of the $i$th instrument at time $t$ when the market conditions are as summarized in $m$. 
Notes:

- $m$ is multivariate: it must contain yield curves and exchange rates for all currencies represented in the portfolio, volatilities of any underlying variable involved in option-like instruments, and equity (stock-market) variables if necessary.

- Market conditions at time $t$ will be denoted $m_t$; thus the value of the $i$th instrument at time $t$ is actually $V_i(t, m_t)$.

- $V_i(t, m)$ depends on time $t$ directly as well as through $m$, because for instance the remaining maturity decreases as $t$ increases.

- $V_i(t, m)$ is a nonlinear function of $m$—mildly so for simple instruments like bonds and swaps, considerably more so for options and option-like instruments, especially when the “exercise date” of the option is near.
P&L distribution

If (and this is a major assumption) no changes are made to the portfolio between times $t$ (representing the “present”) and $t + \delta t$, then the change in value of the portfolio is

$$\delta V = \sum_i V_i(t + \delta t, m_{t+\delta t}) - V_i(t, m_t).$$

The change $\delta V$ represents net profit and loss, or P&L.
Questions about the “level of risk” in the portfolio may be answered by using the probability distribution of $\delta V$ for an appropriate $\delta t$, conditional on $m_u, u \leq t$. Often, the only aspect of this distribution that is used is the “Value at Risk”, namely the (magnitude of the) lower $100(1 - \alpha)\%$ percentile:

$$P(-\delta V \leq \text{VaR}) = \alpha.$$ 

That is, the VaR is the level of loss over $\delta t$ days that will be exceeded with a probability of $1 - \alpha$. Typically, $\delta t$ is one or ten days, and $\alpha$ is .95 or .99. On the assumption, again, that the positions are not changed over the relevant time window, we would expect that on one day out of twenty, the losses would actually exceed the one-day 95% VaR.
Computing The Distribution of P&L

How can we calculate this distribution? An exact analytic approach is almost always infeasible. The simplest approach is:

- approximate \( V_i(t, m) \) by a first order expansion in \( \delta m \) and \( \delta t \);
- approximate the joint (conditional) distribution of \( \delta m \) by a multivariate Gaussian;
- estimate the necessary variances and covariances (means over short times are often assumed to be zero); and
- compute the resulting Gaussian distribution of \( \delta V \), which of course has mean zero and is therefore characterized by its standard deviation.
The value functions $V_i(t, m)$ are rarely given explicitly, whence the required partial derivatives must be obtained numerically, by evaluating $V$ under perturbed market conditions close to $m_t$. The time derivative $\partial V / \partial t$ is typically small enough to be ignored.
This approach is of course subject to the criticism that either or both of the approximations may be too inexact. If either is dropped, the resulting distribution of $\delta V$ becomes nonGaussian. The only approaches that are currently feasible involve simulation:

- draw a sample $m$ from the (conditional) distribution of $m_{t+\delta t}$;
- value every instrument at the sampled $m$; and
- save the resulting value of the portfolio.

After enough iterations, the saved values may be used to estimate any functional of the distribution of $\delta V$. 
Several choices must be made:

- The conditional distribution of $δm$ may be assumed to belong to some parametric family. Alternatively, parametric assumptions may be avoided by bootstrap methods: resampling $δm$ from an empirical distribution of observed changes in past data.
- The valuations may be “exact”, that is, obtained from the same model used to price the instruments, or may be approximations.
- The portfolio valuations may be used directly to obtain empirical percentiles, or a parametric family may be fitted to them; for instance, the left tail of the distribution could be fitted by a Generalized Pareto density:

$$f(x; \xi, \beta) = \frac{1}{\beta} \left(1 - \frac{x}{\beta}\right)^{\xi^{-1}-1}.$$
Measures of Risk

As noted above, for risk management purposes, the P&L distribution is usually summarized by a single number that quantifies the magnitude of the risk, and this risk measure is typically VaR, a quantile.
Artzner, Delbaen, Eber, and Heath (1998; link on the course home page) introduced four criteria that a risk measure should meet, and call a measure that satisfies all four coherent. The concept is also discussed by Duffie and Singleton in Section 2.4.5. The setting is a sample space $\Omega$ of outcomes, and portfolio payoffs (random variables) $X(\omega)$, $Y(\omega)$, ... , and a risk measure $m(X)$ that associates a real-valued risk quantity with any portfolio payoff.
The criteria are:

- **Subadditivity**: For any portfolio payoffs $X$ and $Y$, 
  \[ m(X + Y) \leq m(X) + m(Y); \]
- **Homogeneity**: For any number $\theta > 0$, \[ m(\theta X) = \theta m(X); \]
- **Monotonicity**: \[ m(X) \geq m(Y) \text{ if } X \leq Y; \]
- **Risk-free condition**: \[ m(X + k) = m(X) - k, \text{ for any constant } k. \]
Of these, Homogeneity is debatable: if \( \theta \) is large enough, the risk in the position may become unacceptable; however, it is innocuous for moderate values. Subadditivity is necessary to recognize diversification: combining two portfolios cannot increase the combined risk.

Subadditivity and Homogeneity imply:

- **Convexity:** For any \( 0 \leq \theta \leq 1 \),

  \[
  m(\theta X + (1 - \theta) Y) \leq \theta m(X) + (1 - \theta) m(Y).
  \]

In some recent work, Subadditivity and Homogeneity are replaced by the weaker assumption of Convexity.
Note that no probability measure on $\Omega$ has been introduced: the criteria for coherence do not require one. However, when we have one, using it to propose a candidate $m(\cdot)$ seems reasonable. The formal definition of $\text{VaR}_\alpha$ is

$$\text{VaR}_\alpha(X) = \inf\{x : P(-X > x) \leq 1 - \alpha\}.$$
If $X$ is continuous with a strictly positive density function, $\text{VaR}_\alpha(X)$ is the $\alpha-$quantile of $-X$; that is, the unique solution $v$ of the equation

$$P(-X \leq v) = \alpha.$$ 

The formal definition continues to work when this equation has either no solution or many solutions.
VaR clearly satisfies Homogeneity, Monotonicity, and the Risk-free condition, but fails Subadditivity. To see this, suppose that $X$ and $Y$ are independent and each is $-a < 0$ with probability $p$ and $0$ with probability $1 - p$, and $(1 - \alpha)/2 < p < 1 - \alpha$. Then $\text{VaR}_\alpha(X) = \text{VaR}_\alpha(Y) = 0$, but $\text{VaR}_\alpha(X + Y) = a$. 

More realistically, suppose that $X$ and $Y$ are payoffs on short positions in deep-out-of-the-money options that expire before the end of the period of interest. Then again, both $X$ and $Y$ are zero with high probability, but have possibly long left tails on the negative half-line. If the probability of a zero payout is close enough to 1, VaR is zero for each option separately, but may be positive for their sum.
One simple way to construct a coherent risk measure is to use a family $\mathcal{Q}$ of probability measures, and quantify risk by the worst-case expected loss:

$$m_{\mathcal{Q}}(X) = \sup_{Q \in \mathcal{Q}} E_Q(-X).$$

The general idea is that the distributions in $\mathcal{Q}$ should represent various forms of stress, without deviating too far from what is realistic.

Any such $m_{\mathcal{Q}}(\cdot)$ is easily shown to be coherent. The converse is also true, but harder to prove: for any coherent risk measure $m(\cdot)$ there exists such a family $\mathcal{Q}$. 
If we have a model that provides a distribution $P$, we could take the family $Q$ to be those measures that are boundedly stressed versions of $P$:

$$Q = \left\{ Q : \frac{dQ}{dP} \leq k \right\}$$

for some constant $k > 1$, which we write as $1/(1 - \alpha)$ for some $\alpha$, $0 < \alpha < 1$.

If $A$ is any event, this means that $Q(A) \leq k \times P(A)$; that is, $Q(A)$ might be larger than $P(A)$, but only within limits.
We can write any such \( Q \) as \( dQ(\omega) = \Lambda(\omega)dP(\omega) \), where \( 0 \leq \Lambda(\omega) \leq k \) and

\[
1 = \int dQ(\omega) = \int \Lambda(\omega)dP(\omega),
\]

or in random variable notation, \( E_P(\Lambda) = 1 \). Then we can write

\[
m(X) = \sup_{\Lambda:0 \leq \Lambda \leq k, E(\Lambda)=1} E_P(-\Lambda X).
\]
Clearly, to maximize $E_P(\Lambda X)$ subject to these conditions, we should make $\Lambda$ as large as possible ($k$) where $X$ is large and negative, and zero otherwise. To keep $E_P(\Lambda) = 1$, the set where $\Lambda = k$ must have probability (at most) $1/k = 1 - \alpha$. If $X$ has a continuous distribution with a positive density, this set is just $\{\omega : -X(\omega) \geq \text{VaR}_\alpha(X)\}$, and

$$m(X) = \frac{1}{1 - \alpha} E_P \left[ -X 1_{\{-X \geq \text{VaR}_\alpha(X)\}} \right]$$

$$= E_P(-X | -X \geq \text{VaR}_\alpha(X)). \tag{1}$$

The right hand side of (1) is sometimes called the Conditional Value at Risk, or $\text{CVaR}_\alpha(X)$. 
In general, the worst-case $\Lambda$ may not be unique, but one version is

$$\Lambda = \begin{cases} 
  k & \text{if } -X > \text{VaR}_\alpha(X), \\
  0 & \text{if } -X < \text{VaR}_\alpha(X), \\
  k' & \text{if } -X = \text{VaR}_\alpha(X),
\end{cases}$$

where $0 \leq k' \leq k$ is determined by the requirement that

$$1 = E_P(\Lambda) = k \times P[-X > \text{VaR}_\alpha(X)] + k' \times P[-X = \text{VaR}_\alpha(X)]$$
Hence

\[ k' = \frac{1 - k \times P[-X > \text{VaR}_\alpha(X)]}{P[-X = \text{VaR}_\alpha(X)]} \]

\[ = \frac{1 - \alpha - P[-X > \text{VaR}_\alpha(X)]}{(1 - \alpha) \times P[-X = \text{VaR}_\alpha(X)]} \]

and

\[ m(X) = \frac{1}{1 - \alpha} \left[ E_p \left[ -X 1_{\{-X > \text{VaR}_\alpha(X)\}} \right] + \{1 - \alpha - P[-X > \text{VaR}_\alpha(X)]\} \text{VaR}_\alpha(X) \right] \]
The resulting risk measure $m(X)$ is called the Expected Shortfall $\text{ES}_\alpha(X)$; it is the same as CVaR only if $P[-X \geq \text{VaR}_\alpha(X)] = \alpha$. It may be shown that

$$\text{ES}_\alpha(X) = \frac{1}{1 - \alpha} \int_{\alpha}^{1} \text{VaR}_{\alpha'}(X) d\alpha',$$

whence it is also known as the Average VaR or AVaR$_\alpha$. 
Note that

\[ \text{ES}_\alpha(X) - \text{VaR}_\alpha(X) = EP \left\{ \Lambda \left[ (X - \text{VaR}_\alpha(X)) \right] \right\} \]
\[ = kEP \left\{ \left[ (X - \text{VaR}_\alpha(X)) \right] \mathbb{1}_{\{X > \text{VaR}_\alpha(X)\}} \right\} \]
\[ = kEP \left\{ \max \left[ 0, (X - \text{VaR}_\alpha(X)) \right] \right\} \]
\[ = kEP \left[ \text{amount of loss in excess of } \text{VaR}_\alpha(X) \right]. \]

That is, \( \text{ES}_\alpha(X) \) exceeds \( \text{VaR}_\alpha(X) \) by an amount that is proportional to the expected value of losses in excess of \( \text{VaR}_\alpha(X) \), which is the actuarially fair premium for insurance against those excess losses.
Market Risk Capital Requirements

VaR as described earlier is usually calculated for an institution’s own internal risk management purposes. Since Basel II, supervisors of financial institutions have recognized that the institutions’ capabilities for measuring the degree of risk that they undertake may be better than those of the supervisors themselves.
There has been a corresponding move (by the Bank for International Settlements in Basel as the international coordinator and by some national supervisors in the European Community and the U.S.) to allow each institution to play some role in the calculations of regulatory capital for covering market risk. The move generally refers to the “internal models” developed by the institutions.
Since capital must not be completely consumed by losses except under extraordinary circumstances, the criteria for defining regulatory capital are more stringent than for conventional VaR. The principal differences are:

- the capital requirement is based on potential market losses over a 10-day *holding period*, instead of a single day;
- the pertinent quantile is the 99th percentile of the distribution of losses, instead of the 95th;
- the required capital is the greater of
  - a multiple (between 3 and 4; see below) of the average value of this percentile over the previous 60 days’ calculations; and
  - the current calculation, for the previous day’s portfolio.
The regulators have also restricted the institutions’ freedom as to how distributions are estimated. For instance, volatility *must* be estimated as the unweighted variance of one year’s worth of historical data; more sophisticated methods such as GARCH are prohibited. Also, certain correlations are required to be set equal to 1, for conservatism.
Testing the validity of VaR calculations is of course a normal part of implementing and maintaining such a capability. When a VaR model is to be used as a basis for determining regulatory capital, such testing has to be formalized and codified. One current proposal is based on *backtesting* the calculation over the recent past.
Backtesting proceeds as follows. Take a given portfolio of instruments, say the “current” book. For each of the most recent 250 trading days:

- value the book;
- re-estimate all parameters in the VaR model based on prior data;
- calculate the VaR of the book based on these re-estimated parameters;
- revalue the book one day (or one holding period) later, and subtract to obtain the actual change in value;
- determine whether the change exceeded the VaR (in the direction of loss).

Short Term Variability
The number of occasions on which the VaR was exceeded should be around \(250 \times 0.01 = 2.5\) occasions. The outcomes are classified as follows.

**Green zone:** Finding up to 4 exceedances is viewed as bad luck.

**Yellow zone:** More than 4 but no more than 9 is viewed as requiring an empirical adjustment, which consists of scaling the multiplier of 3 progressively up to 4.

**Red zone:** Finding 10 or more exceedances is bad news.
Incentives

Using VaR to determine minimum regulatory capital faces two criticisms:

- From the technical perspective, VaR is not coherent.
- Practically, it gives a bank an incentive to manage its business so that losses of moderate probability are small, but no incentive to control the magnitude of highly improbable losses.
Suppose that a bank holds capital $C$, and that its one-year P&L is $X$.

The probability that it is insolvent in one year is

$$1 - \alpha = P[(-X) > C],$$

so $C = \text{VaR}_\alpha(X)$.

Its cost of capital is around the insolvency probability, so the cost of holding capital $C$ is

$$(1 - \alpha)C = (1 - \alpha)\text{VaR}_\alpha(X).$$

The bank has no financial incentive to control the magnitude of losses with probability less than $(1 - \alpha)$.
Now suppose that the bank was required to carry insurance against the losses that its creditors would face in an insolvency.

The losses are

\[ L = \begin{cases} \text{if } (-X) > \text{VaR}_\alpha(X) \quad \text{then } & (-X) - \text{VaR}_\alpha(X) \\ 0 & \text{otherwise}. \end{cases} \]

\[ = \max[0, (-X) - \text{VaR}_\alpha(X)]. \]

The expected loss is therefore

\[ E_P \left\{ \max[0, (-X) - \text{VaR}_\alpha(X)] \right\} \]

which is the actuarially fair premium for the insurance.
The combined cost of capital and insurance is

\[(1 - \alpha) \text{VaR}_\alpha(X) + \mathbb{E}_P \{\max[0, (-X) - \text{VaR}_\alpha(X)]\}\]

\[= (1 - \alpha) \text{ES}_\alpha(X).\]

So a bank that is required to pay into an insurance program to cover the losses that it would cause if it became insolvent bears combined capital and insurance costs that are proportional to its expected shortfall.

The bank no longer has an incentive to risk large losses of low probability.
The Basel III documents propose replacing VaR with ES for determining capital requirements.

This proposal would change a bank’s cost structure to reflect the impact of large losses of low probability, in roughly the same way as requiring insurance.

It does not, however, provide a mechanism for compensating society for the losses caused by an insolvency.

Instead of requiring extra capital, Basel III (IV?) should require contributions, based on ES — VAR, to an FDIC-like insurance program.
Allocation of Risk

- When risk in an aggregate of sources has been measured coherently, disaggregating (or *allocating*) it across those sources is often necessary.

- Example: if the risk in a portfolio of investments can be disaggregated into parts associated with each individual investment, they can be compared in terms of risk-adjusted return.
Example: each business unit in a financial institution contributes to the firm’s overall risk.

The firm-wide risk imposes costs, of capital and perhaps of required insurance.

Those costs must be met by a contribution from each business unit, which reduces its profitability.

Determining those contributions amounts to an allocation of firm-wide risk.
If $m(\cdot)$ is a coherent risk measure, an *allocation principle* is a rule for constructing $\{K_i\}$ such that when

$$X = \sum_{i=1}^{n} X_i,$$

the risk is disaggregated as

$$m(X) = \sum_{i=1}^{n} K_i.$$
An allocation principle is coherent if it has three properties:
- No undercut;
- Symmetry;
- Riskless allocation.

Write

\[ N = \{1, 2, \ldots, n\}. \]

The three properties are defined in terms of subsets \( M \subset N \).
- **No undercut**: \( \forall M \subset N, \)

\[
\sum_{i \in M} K_i \leq m \left( \sum_{i \in M} X_i \right).
\]

- In particular, \( \forall i \in N, \)

\[
K_i \leq m(X_i).
\]

- Each business unit receives a smaller allocation than if it were stand-alone.
- No unit has an incentive to break away from the firm.
- More generally, no *coalition* of units has that incentive.
**Symmetry:** if \( \forall M \subset N \setminus \{i, j\} \)

\[
m(M \cup \{i\}) = m(M \cup \{j\}),
\]

then \( K_i = K_j \).

That is, if \( X_i \) and \( X_j \) cause the same incremental risk when added to any subset, they must receive the same allocation.

This makes the allocation purely risk-driven.

**Riskless allocation:** if \( X_n \) is \( k \) units of the risk-free asset, then

\[
K_n = m(X_n) = -k.
\]
Game theory result: under a differentiability condition, the unique coherent allocation principle is

\[ K_j = \frac{\partial}{\partial \lambda} m \left( \sum_{i \neq j} X_i + \lambda X_j \right) \bigg|_{\lambda=1}. \]

That is, the allocation is also the incremental contribution to aggregate risk.
Recall: because $m(\cdot)$ is coherent, there exists a family $Q$ of probability measures such that

$$m(X) = \sup_{Q \in Q} E_Q(-X).$$

Suppose that the supremum is attained at $Q_X$:

$$m(X) = E_{Q_X}(-X)$$

$$\quad = E_{Q_X} \left( \sum -X_i \right)$$

$$\quad = \sum E_{Q_X}(-X_i).$$
This suggests the allocation

$$K_i = E_{Q_x}(-X_i).$$

Is it the *coherent* allocation?

If $m(\cdot)$ is expected shortfall, and $X$ is continuous, yes: the coherent allocation of

$$\text{ES}_\alpha(X) = E_P[-X | -X \geq \text{VaR}_\alpha(X)]$$

is

$$K_i = E_P[-X_i | -X \geq \text{VaR}_\alpha(X)].$$
This allocation of ES is called the *Marginal* Expected Shortfall, MES.

Note that it is easy to calculate in a simulation approach:
- For each simulation, save the total $X$ and the contributions \( \{X_1, X_2, \ldots, X_n\} \);
- Use the simulated totals to find VaR;
- Use the totals exceeding VaR to find ES;
- Use the corresponding contributions to find MES.

MES has been used to measure systemic risk.
Credit Risk and the Use of Collateral

Short term market variability may cause credit risk as well as market risk. As laid out above, market risk is the threat to an institution of losses caused by the impact of adverse market changes on its own portfolio. Credit risk is the indirect threat that adverse market moves may cause losses in another institution’s portfolio, driving that institution into insolvency. If the first institution is exposed to the insolvent party, that is if the insolvency triggers an obligation of the insolvent party to make a payment, then the insolvency may cause economic loss for the first institution.
Many parties now use the pledging of *collateral* to manage this short term credit risk. The party with the obligation to pay delivers securities (Treasury issues, for instance) to the other party. The securities remain the property of the pledgor, who also continues to receive coupon payments, exercise voting rights, and so on. The pledgee holds a *first security interest* in the securities, however, which gives it the right to assume ownership in specified circumstances. *In favorable legal jurisdictions*, this right supersedes the right of a bankruptcy court to control the disposition of the assets of an insolvent entity.
The question arises of how much collateral, and of what kind, constitutes an adequate protection. This question may be answered using concepts from the calculation of Value at Risk. Specifically, suppose that the concerned institution ("us") has a portfolio of $n$ transactions with another entity ("them"), with values $V_i(t, m)$ at time $t$ under market conditions $m$. 
We assume that

\[ V(t, m_t) = \sum_{i=1}^{n} V_i(t, m_t) > 0 \]

(the portfolio is, on a net basis, “in the money” to us at time \(t\)). If “we” hold as collateral \(n'\) securities with values \(V_i'(t, m)\), our unsecured exposure to them is

\[ V(t, m_t) - \sum_{i=1}^{n'} V_i'(t, m_t) = \sum_{i=1}^{n+n'} V_i(t, m_t) = V_U(t, m_t), \]

where we write \(V_{n+i}(\cdot) = -V_i'(\cdot), i = 1, 2, \ldots, n'\).
This the value of a hypothetical portfolio consisting of the existing deals and short positions in the collateral securities. The collateral is satisfactory if the hypothetical portfolio has negative value now and a high probability of staying negative for an acceptable period of time. That is, we want

\[ V_U(t, m_t) \leq 0 \]

and

\[ P(V_U(t + \delta t, m_{t+\delta t}) = V_U(t, m_t) + \delta V \leq 0) \geq 1 - \epsilon. \]

Strictly speaking, we should require

\[ P(V_U(t + u, m_{t+u}) \leq 0, 0 \leq u \leq \delta t) \geq 1 - \epsilon, \]

but the additional complication is not worthwhile.
Note that

\[
P(V_U(t + \delta t, m_{t+\delta t}) = V_U(t, m_t) + \delta V \leq 0) = P(\delta V \leq -V_U(t, m_t))
\]

is essentially the same quantity that we needed to find the VaR. The differences are:

- \(\delta t\) may be longer than one day. One week is common, and one month also occurs. Longer intervals increase the variability of \(\delta V\) and thus increase the *over-collateralization*, but mean less frequent transfers of collateral.

- \(\epsilon\) may not be .05. “We” may need a degree of protection ranging from
  - extreme (\(\epsilon = .001\) or less), if our credit rating is critically dependent on avoiding loss, to
  - relaxed (\(\epsilon = .15\), “1\(\sigma\)”) if we are not in a position to ask for more.
In practice, the market risk in collateral is handled separately from
the market risk in the portfolio that it is pledged to cover. The
conventional approach is to apply a “haircut” to the collateral
to cover possible loss of value.
This is roughly equivalent to assuming that the value of the collateral
has correlation $-1$ with the value of the position to which it is
applied. The conventional approach presumably results in over-collateralization
...but that may not be a bad thing.