Two Tools of a Risk Quant: Time Series Analysis and Value at Risk

Thomas P. Harte

Morgan Stanley[1]
New York

1

Baruch College
The City University of New York
Wednesday, 2015-04-22

[1] Disclaimer: The views expressed in this presentation are those of the author and not Morgan Stanley. The content of this presentation is provided solely for informational purposes and academic discussion at New York University, and it is not intended for use by, and should not be disseminated to, nor relied upon by, any third party in any manner whatsoever. This presentation does not contain anything that constitutes an offer to purchase or sell any instrument nor to enter into, replace, transfer, assign, or terminate any transaction, security, or instrument nor a commitment by Morgan Stanley to make such an offer.

Tools of the Risk Quant

- **Time Series**
  the lifeblood of finance

Source: **TradingView.com**

Tools of the Risk Quant

- **Value at Risk**
  the fundamental risk model

Business **Financial Markets** posted: 2 months ago

JPMorgan Tells SEC New VaR Model Didn't Require Prior Disclosure As Whale Impact Fades

JPMorgan Chase, facing criticism that it misled investors about a change to a risk model as trades backfired last year, told U.S. regulators that the bank wasn't obligated to disclose the move until May.

Bloomberg reports that while there was an 'interim change' to the lender's so-called value-at-risk model during the first three months of 2012, that adjustment had been reversed by the time the company filed its quarterly report in May, then-Chief Financial Officer Douglas Braunstein told the Securities and Exchange Commission in a December 3rd letter that was released Wednesday.

'As a result, the firm believes there was no model change within the meaning of' securities-disclosure laws, he wrote.

Motivation

Motivation

- **Time Series**

Motivation

- **Time Series**
  - (Big words, big words, ...)

Motivation

- **Time Series**
  - Stochastic Process

Motivation

- **Time Series**
  - Stochastic Process
  - Deterministic Process

Motivation

- **Time Series**
  - Stochastic Process
  - Deterministic Process
  - Stationarity

Motivation

- **Time Series**
  - Stochastic Process
  - Deterministic Process
  - Stationarity
  - Non-stationarity

Motivation

- **Time Series**
  - Stochastic Process
  - Deterministic Process
  - Stationarity
  - Non-stationarity
  - Integrated Processes

Motivation

- **Time Series**
  - Stochastic Process
  - Deterministic Process
  - Stationarity
  - Non-stationarity
  - Integrated Processes (and random walks)

Motivation

- **Time Series**
  - Stochastic Process
  - Deterministic Process
  - Stationarity
  - Non-stationarity
  - Integrated Processes (and random walks)
  - Trending

Motivation

- **Time Series**
  - Stochastic Process
  - Deterministic Process
  - Stationarity
  - Non-stationarity
  - Integrated Processes (and random walks)
  - Trending and De-trending

Motivation

- **Time Series**
  - Stochastic Process
  - Deterministic Process
  - Stationarity
  - Non-stationarity
  - Integrated Processes (and random walks)
  - Trending and De-trending
  - Unit Roots

Motivation

- **Time Series**
  - Stochastic Process
  - Deterministic Process
  - Stationarity
  - Non-stationarity
  - Integrated Processes (and random walks)
  - Trending and De-trending
  - Unit Roots
  - Autoregressive Models

Motivation

- **Time Series**
    - Stochastic Process
    - Deterministic Process
    - Stationarity
    - Non-stationarity
    - Integrated Processes (and random walks)
    - Trending and De-trending
    - Unit Roots
    - Autoregressive Models, Moving Average Models

Motivation

- **Time Series**
  - Stochastic Process
  - Deterministic Process
  - Stationarity
  - Non-stationarity
  - Integrated Processes (and random walks)
  - Trending and De-trending
  - Unit Roots
  - Autoregressive Models, Moving Average Models, and combinations thereof…

Motivation

- **Time Series**
  - Stochastic Process
  - Deterministic Process
  - Stationarity
  - Non-stationarity
  - Integrated Processes (and random walks)
  - Trending and De-trending
  - Unit Roots
  - Autoregressive Models, Moving Average Models, and combinations thereof...
- We'll cover the most important aspects:

Motivation

- **Time Series**
  - Stochastic Process
  - Deterministic Process
  - Stationarity
  - Non-stationarity
  - Integrated Processes (and random walks)
  - Trending and De-trending
  - Unit Roots
  - Autoregressive Models, Moving Average Models, and combinations thereof...
- We'll cover the most important aspects:
  - Relate time-series terminology to simple regression models

Motivation

- **Time Series**
  - Stochastic Process
  - Deterministic Process
  - Stationarity
  - Non-stationarity
  - Integrated Processes (and random walks)
  - Trending and De-trending
  - Unit Roots
  - Autoregressive Models, Moving Average Models, and combinations thereof…
- We'll cover the most important aspects:
  - Relate time-series terminology to simple regression models
  - Answer the questions:

Motivation

- **Time Series**
  - Stochastic Process
  - Deterministic Process
  - Stationarity
  - Non-stationarity
  - Integrated Processes (and random walks)
  - Trending and De-trending
  - Unit Roots
  - Autoregressive Models, Moving Average Models, and combinations thereof...
- We'll cover the most important aspects:
  - Relate time-series terminology to simple regression models
  - Answer the questions:
    - What assumptions does the regression model make?

Motivation

- **Time Series**
  - Stochastic Process
  - Deterministic Process
  - Stationarity
  - Non-stationarity
  - Integrated Processes (and random walks)
  - Trending and De-trending
  - Unit Roots
  - Autoregressive Models, Moving Average Models, and combinations thereof...
- We'll cover the most important aspects:
  - Relate time-series terminology to simple regression models
  - Answer the questions:
    - What assumptions does the regression model make?
    - How do we translate time-series problems into a regression model?

Motivation

- **Time Series**
  - Stochastic Process
  - Deterministic Process
  - Stationarity
  - Non-stationarity
  - Integrated Processes (and random walks)
  - Trending and De-trending
  - Unit Roots
  - Autoregressive Models, Moving Average Models, and combinations thereof...
- We'll cover the most important aspects:
  - Relate time-series terminology to simple regression models
  - Answer the questions:
    - What assumptions does the regression model make?
    - How do we translate time-series problems into a regression model?
    - How do we ensure that this translation is valid?

Motivation

Motivation

- **Value at Risk**

Motivation

- **Value at Risk ("VaR")**

Motivation

- **Value at Risk (*never* "VAR")**

Motivation

- **Value at Risk (*never* "VAR", which means "Vector Auto-Regressive")**

Motivation

- **Value at Risk (“VaR”)**
  - P&L is stationary

Motivation

- **Value at Risk ("VaR")**
  - P&L is stationary
  - We can forecast P&L if it's stationary

Motivation

- **Value at Risk ("VaR")**
  - P&L is stationary
  - We can forecast P&L if it's stationary
  - Normal linear VaR

Motivation

- **Value at Risk ("VaR")**
  - P&L is stationary
  - We can forecast P&L if it's stationary
  - Normal linear VaR ("Good old VaR")

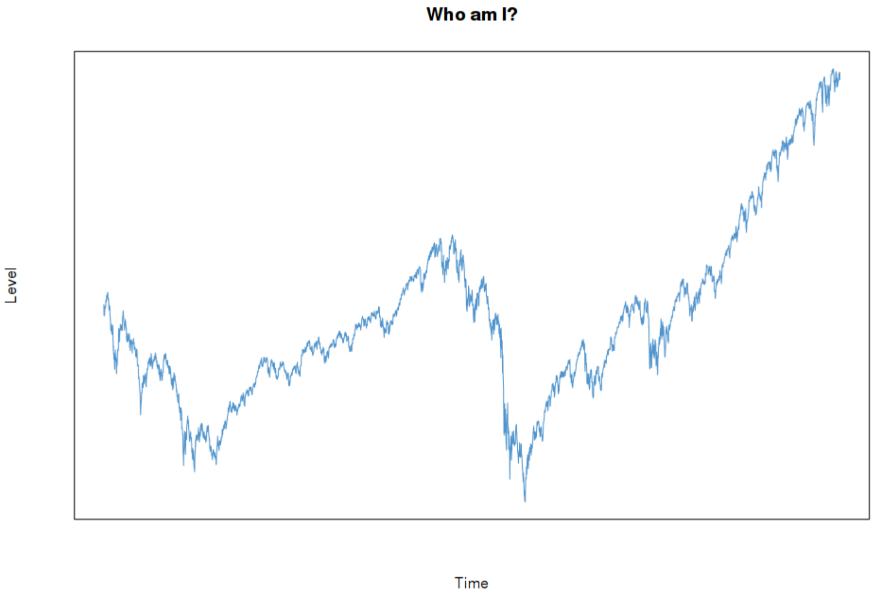
Motivation

- **Value at Risk (“VaR”)**
  - P&L is stationary
  - We can forecast P&L if it’s stationary
  - Normal linear VaR (“Good old VaR”)
  - Why the “Greeks” matter

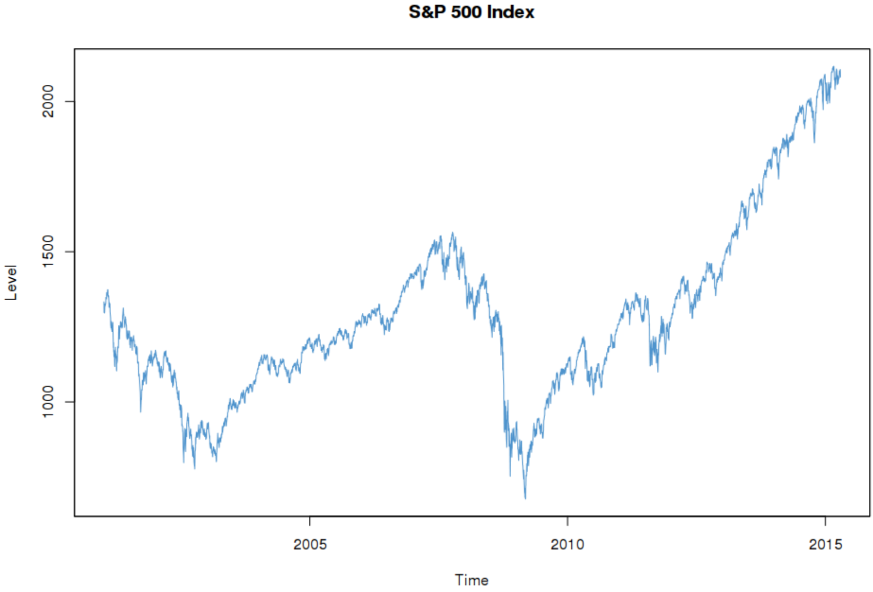
Motivation

- **Value at Risk ("VaR")**
  - P&L is stationary
  - We can forecast P&L if it's stationary
  - Normal linear VaR ("Good old VaR")
  - Why the "Greeks" matter
  - At the end of the lecture we'll relate VaR to time-series analysis through a simple time-series model for the P&L forecast of a FX forward contract

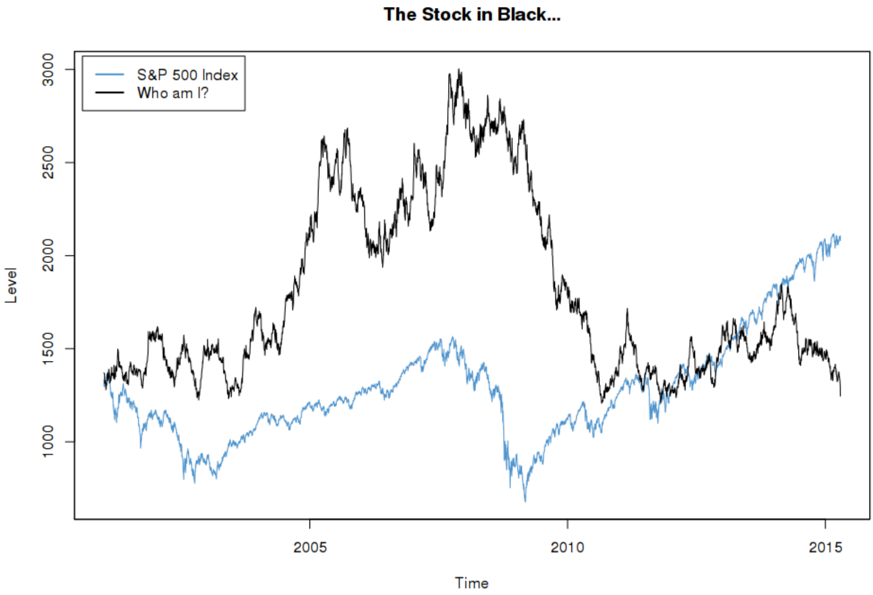
Guess that Stock...



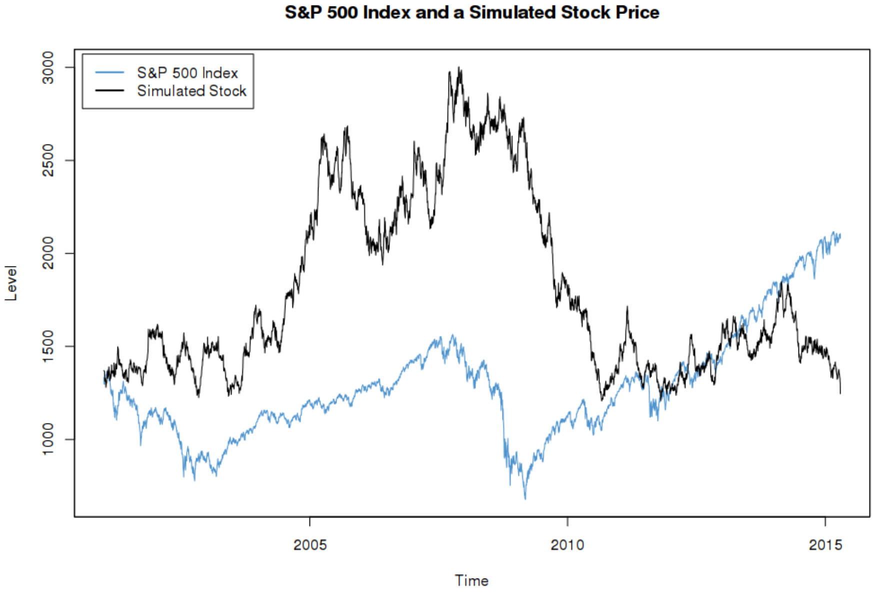
Guess that Stock. . .



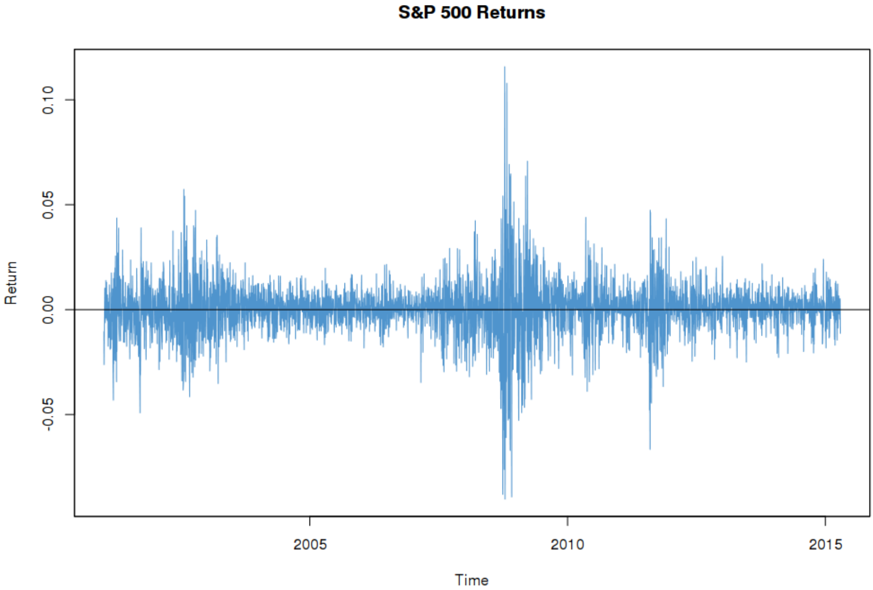
Guess that Stock...



Price Levels are Non-stationary. . .

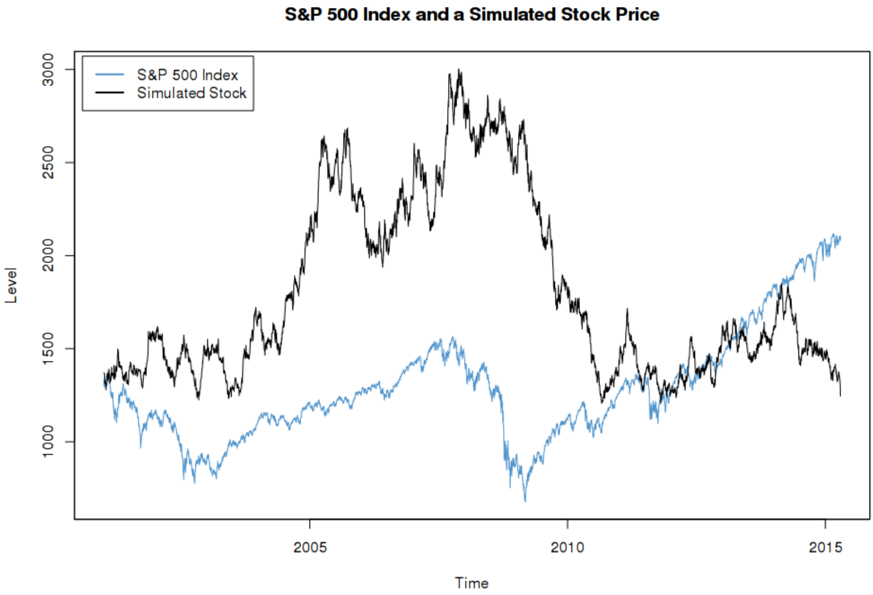


Returns are (mostly) Stationary. . .



Where did the Simulated Stock come from?

Where did the Simulated Stock come from?



Where did the Simulated Stock come from?

- $$\frac{\mathrm{d}S}{S} = r\,\mathrm{d}t + \sigma\,\mathrm{d}W, \tag{1}$$

  where r and σ are assumed to be constant.

Where did the Simulated Stock come from?

- $$\frac{dS}{S} = r\,dt + \sigma\,dW, \tag{1}$$

  where r and σ are assumed to be constant.
- By applying Itō's Lemma and discretizing, Equation (1) becomes

  $$\log P_t = c + \log P_{t-1} + \varepsilon_t, \tag{2}$$

  where $c = r - \sigma^2/2$ and the error term $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$.

Where did the Simulated Stock come from?

- $$\frac{\mathrm{d}S}{S} = r\,\mathrm{d}t + \sigma\,\mathrm{d}W, \tag{1}$$

  where r and σ are assumed to be constant.
- By applying Itō's Lemma and discretizing, Equation (1) becomes

  $$\log P_t = c + \log P_{t-1} + \varepsilon_t, \tag{2}$$

  where $c = r - \sigma^2/2$ and the error term $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$.
- This is the random-walk model that is applied to model (log) prices under the Efficient Markets Hypothesis.

Where did the Simulated Stock come from?

- $$\frac{\mathrm{d}S}{S} = r\,\mathrm{d}t + \sigma\,\mathrm{d}W, \tag{1}$$

  where r and σ are assumed to be constant.
- By applying Itō's Lemma and discretizing, Equation (1) becomes

  $$\log P_t = c + \log P_{t-1} + \varepsilon_t, \tag{2}$$

  where $c = r - \sigma^2/2$ and the error term $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$.
- This is the random-walk model that is applied to model (log) prices under the Efficient Markets Hypothesis. The EMH only requires $\varepsilon_t \overset{\text{i.i.d.}}{\sim} (0, \sigma^2)$

Where did the Simulated Stock come from?

- $$\frac{\mathrm{d}S}{S} = r\,\mathrm{d}t + \sigma\,\mathrm{d}W, \tag{1}$$

  where r and σ are assumed to be constant.
- By applying Itō's Lemma and discretizing, Equation (1) becomes

  $$\log P_t = c + \log P_{t-1} + \varepsilon_t, \tag{2}$$

  where $c = r - \sigma^2/2$ and the error term $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$.
- This is the random-walk model that is applied to model (log) prices under the Efficient Markets Hypothesis. The EMH only requires $\varepsilon_t \overset{\text{i.i.d.}}{\sim} (0, \sigma^2)$
- If $c > 0$ log prices are trending upwards.

Where did the Simulated Stock come from?

- $$\frac{\mathrm{d}S}{S} = r\,\mathrm{d}t + \sigma\,\mathrm{d}W, \tag{1}$$

  where r and σ are assumed to be constant.
- By applying Itō's Lemma and discretizing, Equation (1) becomes

  $$\log P_t = c + \log P_{t-1} + \varepsilon_t, \tag{2}$$

  where $c = r - \sigma^2/2$ and the error term $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$.
- This is the random-walk model that is applied to model (log) prices under the Efficient Markets Hypothesis. The EMH only requires $\varepsilon_t \overset{\text{i.i.d.}}{\sim} (0, \sigma^2)$
- If $c > 0$ log prices are trending upwards.
- If $c < 0$ log prices are trending downwards.

Where did the Simulated Stock come from?

- $$\frac{\mathrm{d}S}{S} = r\,\mathrm{d}t + \sigma\,\mathrm{d}W, \tag{1}$$
  where r and σ are assumed to be constant.
- By applying Itō's Lemma and discretizing, Equation (1) becomes
  $$\log P_t = c + \log P_{t-1} + \varepsilon_t, \tag{2}$$
  where $c = r - \sigma^2/2$ and the error term $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$.
- This is the random-walk model that is applied to model (log) prices under the Efficient Markets Hypothesis. The EMH only requires $\varepsilon_t \overset{\text{i.i.d.}}{\sim} (0, \sigma^2)$
- If $c > 0$ log prices are trending upwards.
- If $c < 0$ log prices are trending downwards.
- No obvious drift upon inspection, but we still say that Equation (2) has a *stochastic trend*.

The Random Walk: An Integrated Process

The Random Walk: An Integrated Process

- The random walk

$$\log P_t = c + \log P_{t-1} + \varepsilon_t, \tag{3}$$

The Random Walk: An Integrated Process

- The random walk

$$\log P_t = c + \log P_{t-1} + \varepsilon_t, \tag{3}$$

is *integrated*.

The Random Walk: An Integrated Process

- The random walk
$$\log P_t = c + \log P_{t-1} + \varepsilon_t, \tag{3}$$
is *integrated*.
- A time series is integrated of order n, written $y_t \sim \mathscr{I}(n)$, if the stochastic part is non-stationary but becomes stationary after differencing a minimum of n times (Alexander, *Market Models*, p. 322).

The Random Walk: An Integrated Process

- The random walk
$$\log P_t = c + \log P_{t-1} + \varepsilon_t, \tag{3}$$
is *integrated*.
- A time series is integrated of order n, written $y_t \sim \mathscr{I}(n)$, if the stochastic part is non-stationary but becomes stationary after differencing a minimum of n times (Alexander, *Market Models*, p. 322).
- A random walk is an example of an integrated process of order 1. Generally, we have $\varepsilon_t \sim \mathscr{I}(0)$.

The Random Walk: An Integrated Process

- The random walk
$$\log P_t = c + \log P_{t-1} + \varepsilon_t, \tag{3}$$
is *integrated*.
- A time series is integrated of order n, written $y_t \sim \mathscr{I}(n)$, if the stochastic part is non-stationary but becomes stationary after differencing a minimum of n times (Alexander, *Market Models*, p. 322).
- A random walk is an example of an integrated process of order 1. Generally, we have $\varepsilon_t \sim \mathscr{I}(0)$.
- To see this, repeatedly substituting for $\log P_{t-1}, \log P_{t-2}, \ldots$ gives

The Random Walk: An Integrated Process

- The random walk
$$\log P_t = c + \log P_{t-1} + \varepsilon_t, \tag{3}$$
is *integrated*.
- A time series is integrated of order n, written $y_t \sim \mathscr{I}(n)$, if the stochastic part is non-stationary but becomes stationary after differencing a minimum of n times (Alexander, *Market Models*, p. 322).
- A random walk is an example of an integrated process of order 1. Generally, we have $\varepsilon_t \sim \mathscr{I}(0)$.
- To see this, repeatedly substituting for $\log P_{t-1}, \log P_{t-2}, \ldots$ gives
$$\log P_t = c_0 + ct + \sum_{t=0}^{t-1} \varepsilon_{t-i}, \tag{4}$$
where $c_0 = \log P_0$.

The Random Walk: An Integrated Process

- The random walk
$$\log P_t = c + \log P_{t-1} + \varepsilon_t, \tag{3}$$
is *integrated*.
- A time series is integrated of order n, written $y_t \sim \mathscr{I}(n)$, if the stochastic part is non-stationary but becomes stationary after differencing a minimum of n times (Alexander, *Market Models*, p. 322).
- A random walk is an example of an integrated process of order 1. Generally, we have $\varepsilon_t \sim \mathscr{I}(0)$.
- To see this, repeatedly substituting for $\log P_{t-1}, \log P_{t-2}, \ldots$ gives
$$\log P_t = c_0 + ct + \sum_{t=0}^{t-1} \varepsilon_{t-i}, \tag{4}$$
where $c_0 = \log P_0$.
- The term $\sum_{t=0}^{t-1} \varepsilon_{t-i}$ is non-stationary

The Random Walk: An Integrated Process

- The random walk
$$\log P_t = c + \log P_{t-1} + \varepsilon_t, \tag{3}$$
is *integrated*.
- A time series is integrated of order n, written $y_t \sim \mathscr{I}(n)$, if the stochastic part is non-stationary but becomes stationary after differencing a minimum of n times (Alexander, *Market Models*, p. 322).
- A random walk is an example of an integrated process of order 1. Generally, we have $\varepsilon_t \sim \mathscr{I}(0)$.
- To see this, repeatedly substituting for $\log P_{t-1}, \log P_{t-2}, \ldots$ gives
$$\log P_t = c_0 + ct + \sum_{t=0}^{t-1} \varepsilon_{t-i}, \tag{4}$$
where $c_0 = \log P_0$.
- The term $\sum_{t=0}^{t-1} \varepsilon_{t-i}$ is non-stationary
- The variance of $\sum_{t=0}^{t-1} \varepsilon_{t-i}$ increases with t.

The Random Walk: An Integrated Process

- The random walk
$$\log P_t = c + \log P_{t-1} + \varepsilon_t, \tag{3}$$
is *integrated*.
- A time series is integrated of order n, written $y_t \sim \mathscr{I}(n)$, if the stochastic part is non-stationary but becomes stationary after differencing a minimum of n times (Alexander, *Market Models*, p. 322).
- A random walk is an example of an integrated process of order 1. Generally, we have $\varepsilon_t \sim \mathscr{I}(0)$.
- To see this, repeatedly substituting for $\log P_{t-1}, \log P_{t-2}, \ldots$ gives
$$\log P_t = c_0 + ct + \sum_{t=0}^{t-1} \varepsilon_{t-i}, \tag{4}$$
where $c_0 = \log P_0$.
- The term $\sum_{t=0}^{t-1} \varepsilon_{t-i}$ is non-stationary
- The variance of $\sum_{t=0}^{t-1} \varepsilon_{t-i}$ increases with t.
- How do we make this stationary?

The Random Walk: An Integrated Process

- The random walk
$$\log P_t = c + \log P_{t-1} + \varepsilon_t, \tag{3}$$
is *integrated*.
- A time series is integrated of order n, written $y_t \sim \mathscr{I}(n)$, if the stochastic part is non-stationary but becomes stationary after differencing a minimum of n times (Alexander, *Market Models*, p. 322).
- A random walk is an example of an integrated process of order 1. Generally, we have $\varepsilon_t \sim \mathscr{I}(0)$.
- To see this, repeatedly substituting for $\log P_{t-1}, \log P_{t-2}, \ldots$ gives
$$\log P_t = c_0 + ct + \sum_{t=0}^{t-1} \varepsilon_{t-i}, \tag{4}$$
where $c_0 = \log P_0$.
- The term $\sum_{t=0}^{t-1} \varepsilon_{t-i}$ is non-stationary
- The variance of $\sum_{t=0}^{t-1} \varepsilon_{t-i}$ increases with t.
- How do we make this stationary?
- By using a first-order difference operator, or “differencing” for short

The Humble AR(1) Model: As Stationary as could be

The Humble AR(1) Model: As Stationary as could be

- The order-1 auto-regressive model is simply

$$y_t = y_{t-1}\alpha + \varepsilon_t, \tag{5}$$

where $\varepsilon_t \overset{\text{i.i.d.}}{\sim} (0, \sigma^2)$

The Humble AR(1) Model: As Stationary as could be

- The order-1 auto-regressive model is simply

$$y_t = y_{t-1}\alpha + \varepsilon_t, \tag{5}$$

  where $\varepsilon_t \overset{\text{i.i.d.}}{\sim} (0, \sigma^2)$
- In the random-walk model, let $y_t = \log P_t - \log P_{t-1}$

The Humble AR(1) Model: As Stationary as could be

- The order-1 auto-regressive model is simply

$$y_t = y_{t-1}\alpha + \varepsilon_t, \tag{5}$$

  where $\varepsilon_t \overset{\text{i.i.d.}}{\sim} (0, \sigma^2)$
- In the random-walk model, let $y_t = \log P_t - \log P_{t-1} \approx \frac{P_t - P_{t-1}}{P_t}$, *i.e.* the return

The Humble AR(1) Model: As Stationary as could be

- The order-1 auto-regressive model is simply

$$y_t = y_{t-1}\alpha + \varepsilon_t, \tag{5}$$

where $\varepsilon_t \overset{\text{i.i.d.}}{\sim} (0, \sigma^2)$

- In the random-walk model, let $y_t = \log P_t - \log P_{t-1} \approx \frac{P_t - P_{t-1}}{P_t}$, *i.e.* the return

$$y_t \quad = \quad y_{t-1}\alpha + \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}\Big(0,\ \sigma^2\Big), \qquad \text{for } t = 1, \ldots, N$$

The Humble AR(1) Model: As Stationary as could be

- The order-1 auto-regressive model is simply

$$y_t = y_{t-1}\alpha + \varepsilon_t, \tag{5}$$

where $\varepsilon_t \overset{\text{i.i.d.}}{\sim} (0, \sigma^2)$

- In the random-walk model, let $y_t = \log P_t - \log P_{t-1} \approx \frac{P_t - P_{t-1}}{P_t}$, *i.e.* the return

$$\begin{aligned} y_t &= y_{t-1}\alpha + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}\left(0,\ \sigma^2\right), \quad \text{for } t = 1, \ldots, N \\ y_t &\sim \mathcal{N}\left(y_{t-1}\alpha,\ \sigma^2\right) \end{aligned}$$

The Humble AR(1) Model: As Stationary as could be

- The order-1 auto-regressive model is simply

$$y_t = y_{t-1}\alpha + \varepsilon_t, \tag{5}$$

where $\varepsilon_t \overset{\text{i.i.d.}}{\sim} (0, \sigma^2)$

- In the random-walk model, let $y_t = \log P_t - \log P_{t-1} \approx \frac{P_t - P_{t-1}}{P_t}$, *i.e.* the return

$$\begin{aligned} y_t &= y_{t-1}\alpha + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}\left(0,\ \sigma^2\right), \quad \text{for } t = 1, \ldots, N \\ y_t &\sim \mathcal{N}\left(y_{t-1}\alpha,\ \sigma^2\right) \end{aligned}$$

- AR(1) model is stable if $\mid \alpha \mid < 1$.

The Humble AR(1) Model: As Stationary as could be

- The order-1 auto-regressive model is simply

$$y_t = y_{t-1}\alpha + \varepsilon_t, \tag{5}$$

  where $\varepsilon_t \overset{\text{i.i.d.}}{\sim} (0, \sigma^2)$

- In the random-walk model, let $y_t = \log P_t - \log P_{t-1} \approx \frac{P_t - P_{t-1}}{P_t}$, *i.e.* the return

$$\begin{aligned} y_t &= y_{t-1}\alpha + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}\left(0,\ \sigma^2\right), \quad \text{for } t = 1, \ldots, N \\ y_t &\sim \mathcal{N}\left(y_{t-1}\alpha,\ \sigma^2\right) \end{aligned}$$

- AR(1) model is stable if $\mid \alpha \mid < 1$.
- If $\mid \alpha \mid < 1$, the AR(1) model is stationary.

The Humble AR(1) Model: As Stationary as could be

- The order-1 auto-regressive model is simply

$$y_t = y_{t-1}\alpha + \varepsilon_t, \tag{5}$$

  where $\varepsilon_t \overset{\text{i.i.d.}}{\sim} (0, \sigma^2)$
- In the random-walk model, let $y_t = \log P_t - \log P_{t-1} \approx \frac{P_t - P_{t-1}}{P_t}$, *i.e.* the return

$$\begin{aligned} y_t &= y_{t-1}\alpha + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}\left(0,\ \sigma^2\right), \quad \text{for } t = 1, \ldots, N \\ y_t &\sim \mathcal{N}\left(y_{t-1}\alpha,\ \sigma^2\right) \end{aligned}$$

- AR(1) model is stable if $\mid \alpha \mid < 1$.
- If $\mid \alpha \mid < 1$, the AR(1) model is stationary.
- **Exercise** — prove that:

The Humble AR(1) Model: As Stationary as could be

- The order-1 auto-regressive model is simply

$$y_t = y_{t-1}\alpha + \varepsilon_t, \tag{5}$$

where $\varepsilon_t \overset{\text{i.i.d.}}{\sim} (0, \sigma^2)$

- In the random-walk model, let $y_t = \log P_t - \log P_{t-1} \approx \frac{P_t - P_{t-1}}{P_t}$, *i.e.* the return

$$\begin{aligned} y_t &= y_{t-1}\alpha + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}\left(0,\ \sigma^2\right), \quad \text{for } t = 1, \ldots, N \\ y_t &\sim \mathcal{N}\left(y_{t-1}\alpha,\ \sigma^2\right) \end{aligned}$$

- AR(1) model is stable if $|\ \alpha\ | < 1$.
- If $|\ \alpha\ | < 1$, the AR(1) model is stationary.
- **Exercise** — prove that:

$$\mathbb{E}[y_t] = 0 \tag{6}$$

The Humble AR(1) Model: As Stationary as could be

- The order-1 auto-regressive model is simply

$$y_t = y_{t-1}\alpha + \varepsilon_t, \tag{5}$$

  where $\varepsilon_t \overset{\text{i.i.d.}}{\sim} (0, \sigma^2)$
- In the random-walk model, let $y_t = \log P_t - \log P_{t-1} \approx \frac{P_t - P_{t-1}}{P_t}$, *i.e.* the return

$$\begin{aligned} y_t &= y_{t-1}\alpha + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}\left(0,\ \sigma^2\right), \quad \text{for } t = 1, \ldots, N \\ y_t &\sim \mathcal{N}\left(y_{t-1}\alpha,\ \sigma^2\right) \end{aligned}$$

- AR(1) model is stable if $|\ \alpha\ | < 1$.
- If $|\ \alpha\ | < 1$, the AR(1) model is stationary.
- **Exercise** — prove that:

$$\mathbb{E}[y_t] = 0 \tag{6}$$

$$\mathbb{E}\left[y_t^2\right] = \frac{\sigma^2}{(1 - \alpha^2)} \tag{7}$$

Why is Stationarity so Important?

Why is Stationarity so Important?

The linear-regression model can be stated as:

$$y = X\beta + \varepsilon, \tag{8}$$

where $y, \varepsilon \in \mathbb{R}^N$, $X \in \mathbb{R}^{N \times p}$ and $\beta \in \mathbb{R}^p$

Why is Stationarity so Important?

The linear-regression model can be stated as:

$$y = X\beta + \varepsilon \overset{\text{think}}{\Leftrightarrow} y_t = y_{t-1}\alpha + \varepsilon_t, \tag{8}$$

where $y, \varepsilon \in \mathbb{R}^N$, $X \in \mathbb{R}^{N\times p}$ and $\beta \in \mathbb{R}^p$. For AR(1), $y \Leftrightarrow y_t, X \Leftrightarrow y_{t-1}, \beta \Leftrightarrow \alpha, p = 1$ and $\varepsilon \Leftrightarrow \varepsilon_t$

Why is Stationarity so Important?

The linear-regression model can be stated as:

$$y = X\beta + \varepsilon \overset{\text{think}}{\Leftrightarrow} y_t = y_{t-1}\alpha + \varepsilon_t, \tag{8}$$

where $y, \varepsilon \in \mathbb{R}^N$, $X \in \mathbb{R}^{N\times p}$ and $\beta \in \mathbb{R}^p$. For AR(1), $y \Leftrightarrow y_t, X \Leftrightarrow y_{t-1}, \beta \Leftrightarrow \alpha, p = 1$ and $\varepsilon \Leftrightarrow \varepsilon_t$

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}, \quad X = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,p} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,p} \\ \vdots & \vdots & & \vdots \\ x_{N,1} & x_{N,2} & \ldots & x_{N,p} \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}. \tag{9}$$

Why is Stationarity so Important?

The linear-regression model can be stated as:

$$y = X\beta + \varepsilon \overset{\text{think}}{\Leftrightarrow} y_t = y_{t-1}\alpha + \varepsilon_t, \tag{8}$$

where $y, \varepsilon \in \mathbb{R}^N$, $X \in \mathbb{R}^{N \times p}$ and $\beta \in \mathbb{R}^p$. For AR(1), $y \Leftrightarrow y_t, X \Leftrightarrow y_{t-1}, \beta \Leftrightarrow \alpha, p = 1$ and $\varepsilon \Leftrightarrow \varepsilon_t$

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}, \quad X = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,p} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,p} \\ \vdots & \vdots & & \vdots \\ x_{N,1} & x_{N,2} & \ldots & x_{N,p} \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}. \tag{9}$$

Linear regression rests on the Gauss-Markov assumptions:

Why is Stationarity so Important?

The linear-regression model can be stated as:

$$y = X\beta + \varepsilon \overset{\text{think}}{\Leftrightarrow} y_t = y_{t-1}\alpha + \varepsilon_t, \tag{8}$$

where $y, \varepsilon \in \mathbb{R}^N$, $X \in \mathbb{R}^{N\times p}$ and $\beta \in \mathbb{R}^p$. For AR(1), $y \Leftrightarrow y_t, X \Leftrightarrow y_{t-1}, \beta \Leftrightarrow \alpha, p = 1$ and $\varepsilon \Leftrightarrow \varepsilon_t$

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}, \quad X = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,p} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,p} \\ \vdots & \vdots & & \vdots \\ x_{N,1} & x_{N,2} & \ldots & x_{N,p} \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}. \tag{9}$$

Linear regression rests on the Gauss-Markov assumptions:

1. $\mathbb{E}[\varepsilon_i] = 0, \quad \forall i$

Why is Stationarity so Important?

The linear-regression model can be stated as:

$$y = X\beta + \varepsilon \overset{\text{think}}{\Leftrightarrow} y_t = y_{t-1}\alpha + \varepsilon_t, \tag{8}$$

where $y, \varepsilon \in \mathbb{R}^N$, $X \in \mathbb{R}^{N\times p}$ and $\beta \in \mathbb{R}^p$. For AR(1), $y \Leftrightarrow y_t, X \Leftrightarrow y_{t-1}, \beta \Leftrightarrow \alpha, p = 1$ and $\varepsilon \Leftrightarrow \varepsilon_t$

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}, \quad X = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,p} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,p} \\ \vdots & \vdots & & \vdots \\ x_{N,1} & x_{N,2} & \dots & x_{N,p} \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}. \tag{9}$$

Linear regression rests on the Gauss-Markov assumptions:

1. $\mathbb{E}[\varepsilon_i] = 0, \quad \forall i$
2. $\mathbb{E}\left[\varepsilon_i^2\right] = \sigma^2 < \infty, \quad \forall i$ [homoscedasticity & finite variance]

Why is Stationarity so Important?

The linear-regression model can be stated as:

$$y = X\beta + \varepsilon \overset{\text{think}}{\Leftrightarrow} y_t = y_{t-1}\alpha + \varepsilon_t, \tag{8}$$

where $y, \varepsilon \in \mathbb{R}^N$, $X \in \mathbb{R}^{N\times p}$ and $\beta \in \mathbb{R}^p$. For AR(1), $y \Leftrightarrow y_t, X \Leftrightarrow y_{t-1}, \beta \Leftrightarrow \alpha, p = 1$ and $\varepsilon \Leftrightarrow \varepsilon_t$

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}, \quad X = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,p} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,p} \\ \vdots & \vdots & & \vdots \\ x_{N,1} & x_{N,2} & \ldots & x_{N,p} \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}. \tag{9}$$

Linear regression rests on the Gauss-Markov assumptions:

1. $\mathbb{E}[\varepsilon_i] = 0, \quad \forall i$
2. $\mathbb{E}\left[\varepsilon_i^2\right] = \sigma^2 < \infty, \quad \forall i$ [homoscedasticity & finite variance]
3. $\mathbb{E}[\varepsilon_i, \varepsilon_j] = 0, \quad \forall i \neq j$ [uncorrelated errors]

Why is Stationarity so Important?

The linear-regression model can be stated as:

$$y = X\beta + \varepsilon \overset{\text{think}}{\Leftrightarrow} y_t = y_{t-1}\alpha + \varepsilon_t, \tag{8}$$

where $y, \varepsilon \in \mathbb{R}^N$, $X \in \mathbb{R}^{N\times p}$ and $\beta \in \mathbb{R}^p$. For AR(1), $y \Leftrightarrow y_t, X \Leftrightarrow y_{t-1}, \beta \Leftrightarrow \alpha, p = 1$ and $\varepsilon \Leftrightarrow \varepsilon_t$

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}, \quad X = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,p} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,p} \\ \vdots & \vdots & & \vdots \\ x_{N,1} & x_{N,2} & \dots & x_{N,p} \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}. \tag{9}$$

Linear regression rests on the Gauss-Markov assumptions:

1. $\mathbb{E}[\varepsilon_i] = 0, \quad \forall i$
2. $\mathbb{E}\left[\varepsilon_i^2\right] = \sigma^2 < \infty, \quad \forall i$ [homoscedasticity & finite variance]
3. $\mathbb{E}[\varepsilon_i, \varepsilon_j] = 0, \quad \forall i \neq j$ [uncorrelated errors]
4. X is a fixed set of numbers

Why is Stationarity so Important?

The linear-regression model can be stated as:

$$y = X\beta + \varepsilon \overset{\text{think}}{\Leftrightarrow} y_t = y_{t-1}\alpha + \varepsilon_t, \tag{8}$$

where $y, \varepsilon \in \mathbb{R}^N$, $X \in \mathbb{R}^{N\times p}$ and $\beta \in \mathbb{R}^p$. For AR(1), $y \Leftrightarrow y_t, X \Leftrightarrow y_{t-1}, \beta \Leftrightarrow \alpha, p = 1$ and $\varepsilon \Leftrightarrow \varepsilon_t$

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}, \quad X = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,p} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,p} \\ \vdots & \vdots & & \vdots \\ x_{N,1} & x_{N,2} & \ldots & x_{N,p} \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}. \tag{9}$$

Linear regression rests on the Gauss-Markov assumptions:

1. $\mathbb{E}[\varepsilon_i] = 0, \quad \forall i$
2. $\mathbb{E}\left[\varepsilon_i^2\right] = \sigma^2 < \infty, \quad \forall i$ [homoscedasticity & finite variance]
3. $\mathbb{E}[\varepsilon_i, \varepsilon_j] = 0, \quad \forall i \neq j$ [uncorrelated errors]
4. X is a fixed set of numbers
5. the rank of X is p.

The Assumptions of Messrs. Gauss & Markov

Linear regression rests on the Gauss-Markov assumptions:

1. $\mathbb{E}[\varepsilon_i] = 0, \quad \forall i$
2. $\mathbb{E}\left[\varepsilon_i^2\right] = \sigma^2 < \infty, \quad \forall i$ [homoscedasticity & finite variance]
3. $\mathbb{E}[\varepsilon_i, \varepsilon_j] = 0, \quad \forall i \neq j$ [uncorrelated errors]
4. X is a fixed set of numbers
5. the rank of X is p.

The Assumptions of Messrs. Gauss & Markov

Linear regression rests on the Gauss-Markov assumptions:

1. $\mathbb{E}[\varepsilon_i] = 0, \quad \forall i$
2. $\mathbb{E}\left[\varepsilon_i^2\right] = \sigma^2 < \infty, \quad \forall i$ [homoscedasticity & finite variance]
3. $\mathbb{E}[\varepsilon_i, \varepsilon_j] = 0, \quad \forall i \neq j$ [uncorrelated errors]
4. X is a fixed set of numbers
5. the rank of X is p.

The first three properties taken together assert that the covariance matrix of the errors ε, which are distributed about zero, has the form:

$$\mathbb{E}[\varepsilon\varepsilon'] = \sigma^2 I, \quad \sigma^2 < \infty \tag{10}$$

The Assumptions of Messrs. Gauss & Markov

Linear regression rests on the Gauss-Markov assumptions:

1. $\mathbb{E}[\varepsilon_i] = 0, \quad \forall i$
2. $\mathbb{E}\left[\varepsilon_i^2\right] = \sigma^2 < \infty, \quad \forall i$ [homoscedasticity & finite variance]
3. $\mathbb{E}[\varepsilon_i, \varepsilon_j] = 0, \quad \forall i \neq j$ [uncorrelated errors]
4. X is a fixed set of numbers
5. the rank of X is p.

The first three properties taken together assert that the covariance matrix of the errors ε, which are distributed about zero, has the form:

$$\mathbb{E}[\varepsilon\varepsilon'] = \sigma^2 I, \qquad \sigma^2 < \infty \tag{10}$$

The fifth assumption concerns identifiability and can be simply thought of as ensuring that the least-squares estimate $\hat{\beta} = (X'X)^{-1}X'y$ can be computed, *viz.* that $(X'X)^{-1}$ exists.

$\hat{\beta}$ — the Regression Estimate

[2] The theoretical estimator assumes that unknown quantities are knowable.

$\hat{\beta}$ — the Regression Estimate

- Linear regression is only as good as the assumptions it makes…

[2] The theoretical estimator assumes that unknown quantities are knowable.

$\hat{\beta}$ — the Regression Estimate

- Linear regression is only as good as the assumptions it makes...
- And the estimator(s) it produces — $\hat{\beta}$.

[2] The theoretical estimator assumes that unknown quantities are knowable.

$\hat{\beta}$ — the Regression Estimate

- Linear regression is only as good as the assumptions it makes…
- And the estimator(s) it produces — $\hat{\beta}$.

The properties of the *theoretical*[2] estimator $\hat{\beta}$ will be now be derived to illustrate where the variables on which it depends come from.

[2] The theoretical estimator assumes that unknown quantities are knowable.

What is an estimate of β?

We select β so as to minimize the quantity

What is an estimate of β?

We select β so as to minimize the quantity

$$\begin{aligned} \phi &= \varepsilon'\varepsilon \qquad (11) \\ &= (y - X\beta)'(y - X\beta), \qquad (12) \end{aligned}$$

from Equation (8), *i.e.* $y = X\beta + \varepsilon$.

What is an estimate of β?

We select β so as to minimize the quantity

$$\phi = \varepsilon'\varepsilon \tag{11}$$

$$= (y - X\beta)'(y - X\beta), \tag{12}$$

from Equation (8), *i.e.* $y = X\beta + \varepsilon$. Multiplying out Equation (12):

$$\phi = y'y - \beta'X'y - y'X\beta + \beta'X'X\beta \tag{13}$$

$$= y'y - 2\beta'X'y + \beta'X'X\beta, \tag{14}$$

where $y'X\beta = \beta'X'y$.

What is an estimate of β?

We select β so as to minimize the quantity

$$\begin{aligned} \phi &= \varepsilon'\varepsilon \qquad (11) \\ &= (y - X\beta)'(y - X\beta), \qquad (12) \end{aligned}$$

from Equation (8), *i.e.* $y = X\beta + \varepsilon$. Multiplying out Equation (12):

$$\begin{aligned} \phi &= y'y - \beta'X'y - y'X\beta + \beta'X'X\beta \qquad (13) \\ &= y'y - 2\beta'X'y + \beta'X'X\beta, \qquad (14) \end{aligned}$$

where $y'X\beta = \beta'X'y$.
Differentiating with respect to β gives:

$$\frac{\partial\phi}{\partial\beta} = -2X'y + 2X'X\beta,$$

What is an estimate of β?

We select β so as to minimize the quantity

$$\begin{aligned} \phi &= \varepsilon'\varepsilon && (11) \\ &= (y - X\beta)'(y - X\beta), && (12) \end{aligned}$$

from Equation (8), *i.e.* $y = X\beta + \varepsilon$. Multiplying out Equation (12):

$$\begin{aligned} \phi &= y'y - \beta'X'y - y'X\beta + \beta'X'X\beta && (13) \\ &= y'y - 2\beta'X'y + \beta'X'X\beta, && (14) \end{aligned}$$

where $y'X\beta = \beta'X'y$.
Differentiating with respect to β gives:

$$\frac{\partial\phi}{\partial\beta} = -2X'y + 2X'X\beta,$$

and equating to zero:

$$-2X'y + 2X'X\beta = 0,$$

we find that

$$\hat{\beta} = (X'X)^{-1}X'y. \tag{15}$$

What is an estimate of β?

We select β so as to minimize the quantity

$$\begin{aligned}\phi &= \varepsilon'\varepsilon && (11)\\ &= (y - X\beta)'(y - X\beta), && (12)\end{aligned}$$

from Equation (8), *i.e.* $y = X\beta + \varepsilon$. Multiplying out Equation (12):

$$\begin{aligned}\phi &= y'y - \beta'X'y - y'X\beta + \beta'X'X\beta && (13)\\ &= y'y - 2\beta'X'y + \beta'X'X\beta, && (14)\end{aligned}$$

where $y'X\beta = \beta'X'y$.
Differentiating with respect to β gives:

$$\frac{\partial\phi}{\partial\beta} = -2X'y + 2X'X\beta,$$

and equating to zero:

$$-2X'y + 2X'X\beta = 0,$$

we find that

$$\hat{\beta} = (X'X)^{-1}X'y. \qquad (15)$$

The matrix $X'X$ is non-singular, as guaranteed by Assumption (5), above.

Expected Value of $\hat{\beta}$

As

$$y = X\beta + \varepsilon,$$

Expected Value of $\hat{\beta}$

As

$$y = X\beta + \varepsilon,$$

and the least-squares estimate is:

$$\hat{\beta} = (X'X)^{-1}X'y,$$

Expected Value of $\hat{\beta}$

As

$$y = X\beta + \varepsilon,$$

and the least-squares estimate is:

$$\hat{\beta} = (X'X)^{-1}X'y,$$

we can write $\hat{\beta}$ as

$$\begin{aligned}
\hat{\beta} &= (X'X)^{-1}X'y \\
&= (X'X)^{-1}X'(X\beta + \varepsilon) \\
&= (X'X)^{-1}X'X\beta + (X'X)^{-1}X'\varepsilon \\
&= \beta + (X'X)^{-1}X'\varepsilon.
\end{aligned}$$

Expected Value of $\hat{\beta}$

As

$$y = X\beta + \varepsilon,$$

and the least-squares estimate is:

$$\hat{\beta} = (X'X)^{-1}X'y,$$

we can write $\hat{\beta}$ as

$$\begin{aligned}
\hat{\beta} &= (X'X)^{-1}X'y \\
&= (X'X)^{-1}X'(X\beta + \varepsilon) \\
&= (X'X)^{-1}X'X\beta + (X'X)^{-1}X'\varepsilon \\
&= \beta + (X'X)^{-1}X'\varepsilon.
\end{aligned}$$

Taking the expectation, we have

Expected Value of $\hat{\beta}$

As

$$y = X\beta + \varepsilon,$$

and the least-squares estimate is:

$$\hat{\beta} = (X'X)^{-1}X'y,$$

we can write $\hat{\beta}$ as

$$\begin{aligned}\hat{\beta} &= (X'X)^{-1}X'y \\ &= (X'X)^{-1}X'(X\beta + \varepsilon) \\ &= (X'X)^{-1}X'X\beta + (X'X)^{-1}X'\varepsilon \\ &= \beta + (X'X)^{-1}X'\varepsilon.\end{aligned}$$

Taking the expectation, we have

$$\begin{aligned}\mathbb{E}\left[\hat{\beta}\right] &= \mathbb{E}\left[\beta + (X'X)^{-1}X'\varepsilon\right] \\ &= \beta + \mathbb{E}\left[(X'X)^{-1}X'\varepsilon\right] \\ &= \beta + (X'X)^{-1}X'\mathbb{E}[\varepsilon] \\ &= \beta + 0 \\ &= \beta,\end{aligned}$$

Expected Value of $\hat{\beta}$

As

$$y = X\beta + \varepsilon,$$

and the least-squares estimate is:

$$\hat{\beta} = (X'X)^{-1}X'y,$$

we can write $\hat{\beta}$ as

$$\begin{aligned}
\hat{\beta} &= (X'X)^{-1}X'y \\
&= (X'X)^{-1}X'(X\beta + \varepsilon) \\
&= (X'X)^{-1}X'X\beta + (X'X)^{-1}X'\varepsilon \\
&= \beta + (X'X)^{-1}X'\varepsilon.
\end{aligned}$$

Taking the expectation, we have

$$\mathbb{E}\left[\hat{\beta}\right] = \beta,$$

where $\mathbb{E}[\varepsilon] = 0$ came from Assumption (1), above.

Expected Value of $\hat{\beta}$

As

$$y = X\beta + \varepsilon,$$

and the least-squares estimate is:

$$\hat{\beta} = (X'X)^{-1}X'y,$$

we can write $\hat{\beta}$ as

$$\begin{aligned}
\hat{\beta} &= (X'X)^{-1}X'y \\
&= (X'X)^{-1}X'(X\beta + \varepsilon) \\
&= (X'X)^{-1}X'X\beta + (X'X)^{-1}X'\varepsilon \\
&= \beta + (X'X)^{-1}X'\varepsilon.
\end{aligned}$$

Taking the expectation, we have

$$\mathbb{E}\left[\hat{\beta}\right] = \beta,$$

where $\mathbb{E}[\varepsilon] = 0$ came from Assumption (1), above. Thus, the least-squares estimate $\hat{\beta}$ is unbiased.

Covariance of $\hat{\beta}$

We have just seen that

$$\hat{\beta} = \beta + (X'X)^{-1}X'\varepsilon. \tag{16}$$

Covariance of $\hat{\beta}$

We have just seen that

$$\hat{\beta} = \beta + (X'X)^{-1}X'\varepsilon. \tag{16}$$

Thus, taking the covariance of $\hat{\beta}$ as

$$V(\hat{\beta}) = \mathbb{E}\Big[(\hat{\beta}-\beta)(\hat{\beta}-\beta)'\Big], \tag{17}$$

Covariance of $\hat{\beta}$

We have just seen that

$$\hat{\beta} = \beta + (X'X)^{-1}X'\varepsilon. \tag{16}$$

Thus, taking the covariance of $\hat{\beta}$ as

$$V(\hat{\beta}) = \mathbb{E}\Big[(\hat{\beta}-\beta)(\hat{\beta}-\beta)'\Big], \tag{17}$$

we can re-write $\hat{\beta}-\beta$ as:

$$\hat{\beta}-\beta = (X'X)^{-1}X'\varepsilon, \tag{18}$$

Covariance of $\hat{\beta}$

We have just seen that

$$\hat{\beta} = \beta + (X'X)^{-1}X'\varepsilon. \tag{16}$$

Thus, taking the covariance of $\hat{\beta}$ as

$$V(\hat{\beta}) = \mathbb{E}\Big[(\hat{\beta}-\beta)(\hat{\beta}-\beta)'\Big], \tag{17}$$

we can re-write $\hat{\beta}-\beta$ as:

$$\hat{\beta}-\beta = (X'X)^{-1}X'\varepsilon, \tag{18}$$

and so

$$\begin{aligned}
V(\hat{\beta}) &= \mathbb{E}\Big[(\hat{\beta}-\beta)(\hat{\beta}-\beta)'\Big] \\
&= \mathbb{E}\Big[((X'X)^{-1}X'\varepsilon)((X'X)^{-1}X'\varepsilon)'\Big] \\
&= \mathbb{E}\Big[((X'X)^{-1}X'\varepsilon)(\varepsilon'X(X'X)^{-1})\Big] \\
&= (X'X)^{-1}X'\mathbb{E}[\varepsilon\varepsilon']X(X'X)^{-1} \\
&= (X'X)^{-1}X'(\sigma^2 I)X(X'X)^{-1} \\
&= \sigma^2(X'X)^{-1}X'X(X'X)^{-1} \\
&= \sigma^2(X'X)^{-1}.
\end{aligned} \tag{19}$$

Covariance of $\hat{\beta}$

We have just seen that

$$\hat{\beta} = \beta + (X'X)^{-1}X'\varepsilon. \tag{16}$$

Thus, taking the covariance of $\hat{\beta}$ as

$$V(\hat{\beta}) = \mathbb{E}\Big[(\hat{\beta}-\beta)(\hat{\beta}-\beta)'\Big], \tag{17}$$

we can re-write $\hat{\beta}-\beta$ as:

$$\hat{\beta}-\beta = (X'X)^{-1}X'\varepsilon, \tag{18}$$

and so

$$V(\hat{\beta}) = \sigma^2(X'X)^{-1}.$$

Theory versus Practice

- So far the properties of $\hat{\beta}$ are *theoretical* estimates, where it is assumed that what in practice are unknown quantities are, in fact, knowable.

Theory versus Practice

- So far the properties of $\hat{\beta}$ are *theoretical* estimates, where it is assumed that what in practice are unknown quantities are, in fact, knowable.
- What, therefore, is a a practically realizable estimator $\hat{V}(\hat{\beta})$ for the theoretical covariance of $\hat{\beta}$, $V(\hat{\beta})$?

Theory versus Practice

- So far the properties of $\hat{\beta}$ are *theoretical* estimates, where it is assumed that what in practice are unknown quantities are, in fact, knowable.
- What, therefore, is a a practically realizable estimator $\hat{V}(\hat{\beta})$ for the theoretical covariance of $\hat{\beta}$, $V(\hat{\beta})$?
- We can find an estimator $\hat{V}(\hat{\beta})$ by first focusing on the vector of least-squares residuals:

Theory versus Practice

- So far the properties of $\hat{\beta}$ are *theoretical* estimates, where it is assumed that what in practice are unknown quantities are, in fact, knowable.
- What, therefore, is a a practically realizable estimator $\hat{V}(\hat{\beta})$ for the theoretical covariance of $\hat{\beta}$, $V(\hat{\beta})$?
- We can find an estimator $\hat{V}(\hat{\beta})$ by first focusing on the vector of least-squares residuals:
$$\hat{\varepsilon} = y - X\hat{\beta}. \tag{19}$$

Theory versus Practice

- So far the properties of $\hat{\beta}$ are *theoretical* estimates, where it is assumed that what in practice are unknown quantities are, in fact, knowable.
- What, therefore, is a a practically realizable estimator $\hat{V}(\hat{\beta})$ for the theoretical covariance of $\hat{\beta}$, $V(\hat{\beta})$?
- We can find an estimator $\hat{V}(\hat{\beta})$ by first focusing on the vector of least-squares residuals:

$$\hat{\varepsilon} = y - X\hat{\beta}. \tag{19}$$

- The residuals $\hat{\varepsilon}$ (as opposed to the error term ε) are realizable, as y and X are our data, and are thus known, and we have already seen that $\hat{\beta} = (X'X)^{-1}X'y$, which is computed from our data.

Theory versus Practice

- So far the properties of $\hat{\beta}$ are *theoretical* estimates, where it is assumed that what in practice are unknown quantities are, in fact, knowable.
- What, therefore, is a a practically realizable estimator $\hat{V}(\hat{\beta})$ for the theoretical covariance of $\hat{\beta}$, $V(\hat{\beta})$?
- We can find an estimator $\hat{V}(\hat{\beta})$ by first focusing on the vector of least-squares residuals:

$$\hat{\varepsilon} = y - X\hat{\beta}. \tag{19}$$

- The residuals $\hat{\varepsilon}$ (as opposed to the error term ε) are realizable, as y and X are our data, and are thus known, and we have already seen that $\hat{\beta} = (X'X)^{-1}X'y$, which is computed from our data.
- From Assumption (2), above, we have $\mathbb{E}\left[\varepsilon^2\right] = \sigma^2$.

Theory versus Practice

- So far the properties of $\hat{\beta}$ are *theoretical* estimates, where it is assumed that what in practice are unknown quantities are, in fact, knowable.
- What, therefore, is a a practically realizable estimator $\hat{V}(\hat{\beta})$ for the theoretical covariance of $\hat{\beta}$, $V(\hat{\beta})$?
- We can find an estimator $\hat{V}(\hat{\beta})$ by first focusing on the vector of least-squares residuals:

$$\hat{\varepsilon} = y - X\hat{\beta}. \tag{19}$$

- The residuals $\hat{\varepsilon}$ (as opposed to the error term ε) are realizable, as y and X are our data, and are thus known, and we have already seen that $\hat{\beta} = (X'X)^{-1}X'y$, which is computed from our data.
- From Assumption (2), above, we have $\mathbb{E}\left[\varepsilon^2\right] = \sigma^2$.
- Recall that $V(\hat{\beta}) = \sigma^2(X'X)^{-1}$, where σ^2 is the *theoretical* variance of the error term ε.

Theory versus Practice (continued...)

- Note that $\hat{\varepsilon}'\hat{\varepsilon} \in \mathbb{R}^1$, *i.e.* $\hat{\varepsilon}'\hat{\varepsilon}$ is a scalar.

Theory versus Practice (continued...)

- Note that $\hat{\varepsilon}'\hat{\varepsilon} \in \mathbb{R}^1$, *i.e.* $\hat{\varepsilon}'\hat{\varepsilon}$ is a scalar. Also note that $\hat{\varepsilon}'\hat{\varepsilon}$ is simply the sum of squared residuals:

$$\hat{\varepsilon}'\hat{\varepsilon} = \sum_{i=1}^{n} \hat{\varepsilon}_i^2. \tag{20}$$

Theory versus Practice (continued...)

- Note that $\hat{\varepsilon}'\hat{\varepsilon} \in \mathbb{R}^1$, *i.e.* $\hat{\varepsilon}'\hat{\varepsilon}$ is a scalar. Also note that $\hat{\varepsilon}'\hat{\varepsilon}$ is simply the sum of squared residuals:

$$\hat{\varepsilon}'\hat{\varepsilon} = \sum_{i=1}^{n} \hat{\varepsilon}_i^2. \tag{20}$$

- From Equation (16) we have that $\hat{\beta} = \beta + (X'X)^{-1}X'\varepsilon$, so we can re-express Equation (19) as

Theory versus Practice (continued...)

- Note that $\hat{\varepsilon}'\hat{\varepsilon} \in \mathbb{R}^1$, *i.e.* $\hat{\varepsilon}'\hat{\varepsilon}$ is a scalar. Also note that $\hat{\varepsilon}'\hat{\varepsilon}$ is simply the sum of squared residuals:
$$\hat{\varepsilon}'\hat{\varepsilon} = \sum_{i=1}^{n} \hat{\varepsilon}_i^2. \tag{20}$$
- From Equation (16) we have that $\hat{\beta} = \beta + (X'X)^{-1}X'\varepsilon$, so we can re-express Equation (19) as
$$\hat{\varepsilon} = y - X\hat{\beta}$$

Theory versus Practice (continued...)

- Note that $\hat{\varepsilon}'\hat{\varepsilon} \in \mathbb{R}^1$, *i.e.* $\hat{\varepsilon}'\hat{\varepsilon}$ is a scalar. Also note that $\hat{\varepsilon}'\hat{\varepsilon}$ is simply the sum of squared residuals:

$$\hat{\varepsilon}'\hat{\varepsilon} = \sum_{i=1}^{n} \hat{\varepsilon}_i^2. \tag{20}$$

- From Equation (16) we have that $\hat{\beta} = \beta + (X'X)^{-1}X'\varepsilon$, so we can re-express Equation (19) as

$$\hat{\varepsilon} = y - X\hat{\beta} = (I - X(X'X)^{-1}X')\varepsilon$$

Theory versus Practice (continued...)

- Note that $\hat{\varepsilon}'\hat{\varepsilon} \in \mathbb{R}^1$, *i.e.* $\hat{\varepsilon}'\hat{\varepsilon}$ is a scalar. Also note that $\hat{\varepsilon}'\hat{\varepsilon}$ is simply the sum of squared residuals:

$$\hat{\varepsilon}'\hat{\varepsilon} = \sum_{i=1}^{n} \hat{\varepsilon}_i^2. \tag{20}$$

- From Equation (16) we have that $\hat{\beta} = \beta + (X'X)^{-1}X'\varepsilon$, so we can re-express Equation (19) as

$$\hat{\varepsilon} = y - X\hat{\beta} = (I - X(X'X)^{-1}X')\varepsilon$$

- Assumption (2) has it that $\mathbb{E}\left[\varepsilon_i^2\right] = \sigma^2$ and using the fact that

$$\mathbb{E}[\varepsilon'\varepsilon] = \sum_{i=1}^{N} \mathbb{E}\left[\varepsilon_i^2\right]$$

Theory versus Practice (continued...)

- Note that $\hat{\varepsilon}'\hat{\varepsilon} \in \mathbb{R}^1$, *i.e.* $\hat{\varepsilon}'\hat{\varepsilon}$ is a scalar. Also note that $\hat{\varepsilon}'\hat{\varepsilon}$ is simply the sum of squared residuals:

$$\hat{\varepsilon}'\hat{\varepsilon} = \sum_{i=1}^{n} \hat{\varepsilon}_i^2. \tag{20}$$

- From Equation (16) we have that $\hat{\beta} = \beta + (X'X)^{-1}X'\varepsilon$, so we can re-express Equation (19) as

$$\hat{\varepsilon} = y - X\hat{\beta} = (I - X(X'X)^{-1}X')\varepsilon$$

- Assumption (2) has it that $\mathbb{E}\left[\varepsilon_i^2\right] = \sigma^2$ and using the fact that

$$\mathbb{E}[\varepsilon'\varepsilon] = \sum_{i=1}^{N} \mathbb{E}\left[\varepsilon_i^2\right] = N\sigma^2$$

Theory versus Practice (continued...)

- Note that $\hat{\varepsilon}'\hat{\varepsilon} \in \mathbb{R}^1$, *i.e.* $\hat{\varepsilon}'\hat{\varepsilon}$ is a scalar. Also note that $\hat{\varepsilon}'\hat{\varepsilon}$ is simply the sum of squared residuals:
$$\hat{\varepsilon}'\hat{\varepsilon} = \sum_{i=1}^{n} \hat{\varepsilon}_i^2. \tag{20}$$
- From Equation (16) we have that $\hat{\beta} = \beta + (X'X)^{-1}X'\varepsilon$, so we can re-express Equation (19) as
$$\hat{\varepsilon} = y - X\hat{\beta} = (I - X(X'X)^{-1}X')\varepsilon$$
- Assumption (2) has it that $\mathbb{E}\left[\varepsilon_i^2\right] = \sigma^2$ and using the fact that
$$\mathbb{E}[\varepsilon'\varepsilon] = \sum_{i=1}^{N} \mathbb{E}\left[\varepsilon_i^2\right] = N\sigma^2 \ldots \tag{21}$$

Theory versus Practice (continued again. . .)

Theory versus Practice (continued again...)

- We can write

Theory versus Practice (continued again...)

- We can write
$$\mathbb{E}[\hat{\varepsilon}'\hat{\varepsilon}] = \mathbb{E}\Big[\varepsilon'\varepsilon - \varepsilon' X (X'X)^{-1} X'\varepsilon\Big]$$

Theory versus Practice (continued again...)

- We can write

$$\mathbb{E}[\hat{\varepsilon}'\hat{\varepsilon}] = \mathbb{E}\Big[\varepsilon'\varepsilon - \varepsilon' X(X'X)^{-1}X'\varepsilon\Big] \qquad \text{[after lots of manipulation!]}$$

Theory versus Practice (continued again...)

- We can write

$$\mathbb{E}[\hat{\varepsilon}'\hat{\varepsilon}] = \mathbb{E}\Big[\varepsilon'\varepsilon - \varepsilon' X(X'X)^{-1}X'\varepsilon\Big] = (N-p)\sigma^2$$

Theory versus Practice (continued again...)

- We can write
$$\mathbb{E}[\hat{\varepsilon}'\hat{\varepsilon}] = \mathbb{E}\Big[\varepsilon'\varepsilon - \varepsilon' X(X'X)^{-1}X'\varepsilon\Big] = (N-p)\sigma^2$$

- So therefore we take
$$\hat{\sigma}^2 = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{(N-p)} \tag{22}$$
as an estimator of the error variance.

Theory versus Practice (continued again...)

- We can write
$$\mathbb{E}[\hat{\varepsilon}'\hat{\varepsilon}] = \mathbb{E}\Big[\varepsilon'\varepsilon - \varepsilon' X(X'X)^{-1}X'\varepsilon\Big] = (N-p)\sigma^2$$
- So therefore we take
$$\hat{\sigma}^2 = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{(N-p)} \tag{22}$$
as an estimator of the error variance.
- Combining Equation (22) with Equation (20),

Theory versus Practice (continued again...)

- We can write

$$\mathbb{E}[\hat{\varepsilon}'\hat{\varepsilon}] = \mathbb{E}\Big[\varepsilon'\varepsilon - \varepsilon' X(X'X)^{-1}X'\varepsilon\Big] = (N-p)\sigma^2$$

- So therefore we take

$$\hat{\sigma}^2 = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{(N-p)} \tag{22}$$

  as an estimator of the error variance.

- Combining Equation (22) with Equation (20), *i.e.*

$$V(\hat{\beta}) = \sigma^2(X'X)^{-1}.$$

Theory versus Practice (continued again...)

- We can write
$$\mathbb{E}[\hat{\varepsilon}'\hat{\varepsilon}] = \mathbb{E}\Big[\varepsilon'\varepsilon - \varepsilon' X(X'X)^{-1}X'\varepsilon\Big] = (N-p)\sigma^2$$
- So therefore we take
$$\hat{\sigma}^2 = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{(N-p)} \tag{22}$$
as an estimator of the error variance.
- Combining Equation (22) with Equation (20), *i.e.*
$$V(\hat{\beta}) = \sigma^2(X'X)^{-1}.$$
we have

Theory versus Practice (continued again...)

- We can write

$$\mathbb{E}[\hat{\varepsilon}'\hat{\varepsilon}] = \mathbb{E}\Big[\varepsilon'\varepsilon - \varepsilon' X(X'X)^{-1}X'\varepsilon\Big] = (N-p)\sigma^2$$

- So therefore we take

$$\hat{\sigma}^2 = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{(N-p)} \tag{22}$$

  as an estimator of the error variance.

- Combining Equation (22) with Equation (20), *i.e.*

$$V(\hat{\beta}) = \sigma^2(X'X)^{-1}.$$

  we have

$$\hat{V}(\hat{\beta}) = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{(N-p)}(X'X)^{-1} \tag{23}$$

Theory versus Practice (continued again...)

- We can write

$$\mathbb{E}[\hat{\varepsilon}'\hat{\varepsilon}] = \mathbb{E}\Big[\varepsilon'\varepsilon - \varepsilon' X(X'X)^{-1}X'\varepsilon\Big] = (N-p)\sigma^2$$

- So therefore we take

$$\hat{\sigma}^2 = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{(N-p)} \tag{22}$$

  as an estimator of the error variance.
- Combining Equation (22) with Equation (20), *i.e.*

$$V(\hat{\beta}) = \sigma^2(X'X)^{-1}.$$

  we have

$$\hat{V}(\hat{\beta}) = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{(N-p)}(X'X)^{-1} \tag{23}$$

- All of the variables in Equation (23) we have in practice.

Value at Risk ("VaR")

[3] *i.e.* $100 \times (1 - \alpha) = 99\%$

Value at Risk ("VaR")

- VaR is a loss figure associated with a portfolio

[3] *i.e.* $100 \times (1 - \alpha) = 99\%$

Value at Risk ("VaR")

- VaR is a loss figure associated with a portfolio
- VaR has a holding period

[3] *i.e.* $100 \times (1 - \alpha) = 99\%$

Value at Risk (“VaR”)

- VaR is a loss figure associated with a portfolio
- VaR has a holding period
- VaR has a specified confidence

[3] *i.e.* $100 \times (1 - \alpha) = 99\%$

Value at Risk ("VaR")

- VaR is a loss figure associated with a portfolio
- VaR has a holding period
- VaR has a specified confidence
- We are confident that the loss won't be exceeded if the current portfolio is held for a specified period of time

[3] *i.e.* $100 \times (1 - \alpha) = 99\%$

Value at Risk ("VaR")

- VaR is a loss figure associated with a portfolio
- VaR has a holding period
- VaR has a specified confidence
- We are confident that the loss won't be exceeded if the current portfolio is held for a specified period of time
- VaR is *always* quoted explicitly in terms of two parameters:

[3] *i.e.* $100 \times (1 - \alpha) = 99\%$

Value at Risk ("VaR")

- VaR is a loss figure associated with a portfolio
- VaR has a holding period
- VaR has a specified confidence
- We are confident that the loss won't be exceeded if the current portfolio is held for a specified period of time
- VaR is *always* quoted explicitly in terms of two parameters:
    1. The confidence level $1 - \alpha$, where α is the significance level;

[3] *i.e.* $100 \times (1 - \alpha) = 99\%$

Value at Risk (“VaR”)

- VaR is a loss figure associated with a portfolio
- VaR has a holding period
- VaR has a specified confidence
- We are confident that the loss won’t be exceeded if the current portfolio is held for a specified period of time
- VaR is *always* quoted explicitly in terms of two parameters:
  1. The confidence level $1 - \alpha$, where α is the significance level;
  2. The holding period (or risk horizon) h over which VaR is measured.

[3] *i.e.* $100 \times (1 - \alpha) = 99\%$

Value at Risk ("VaR")

- VaR is a loss figure associated with a portfolio
- VaR has a holding period
- VaR has a specified confidence
- We are confident that the loss won't be exceeded if the current portfolio is held for a specified period of time
- VaR is *always* quoted explicitly in terms of two parameters:
  1. The confidence level $1-\alpha$, where α is the significance level;
  2. The holding period (or risk horizon) h over which VaR is measured.
- We might say that we are 99% certain[3] at the $100 \times \alpha = 1\%$ level of significance, that a loss of \$VaR will not be exceeded in a fund's portfolio over a holding period of $h = 1$ month.

[3] *i.e.* $100 \times (1-\alpha) = 99\%$

Value at Risk ("VaR")

- VaR is a loss figure associated with a portfolio
- VaR has a holding period
- VaR has a specified confidence
- We are confident that the loss won't be exceeded if the current portfolio is held for a specified period of time
- VaR is *always* quoted explicitly in terms of two parameters:
  1. The confidence level $1 - \alpha$, where α is the significance level;
  2. The holding period (or risk horizon) h over which VaR is measured.
- We might say that we are 99% certain[3] at the $100 \times \alpha = 1\%$ level of significance, that a loss of \$VaR will not be exceeded in a fund's portfolio over a holding period of $h = 1$ month.
- VaR is always a forward-looking measurement. . .

[3] *i.e.* $100 \times (1 - \alpha) = 99\%$

Value at Risk ("VaR")

- VaR is a loss figure associated with a portfolio
- VaR has a holding period
- VaR has a specified confidence
- We are confident that the loss won't be exceeded if the current portfolio is held for a specified period of time
- VaR is *always* quoted explicitly in terms of two parameters:
    1. The confidence level $1 - \alpha$, where α is the significance level;
    2. The holding period (or risk horizon) h over which VaR is measured.
- We might say that we are 99% certain[3] at the $100 \times \alpha = 1\%$ level of significance, that a loss of $VaR will not be exceeded in a fund's portfolio over a holding period of $h = 1$ month.
- VaR is always a forward-looking measurement...
- The value that is forecast is of theoretical (or unrealized) P&L

[3] *i.e.* $100 \times (1 - \alpha) = 99\%$

Value at Risk ("VaR"), continued. . .

Value at Risk ("VaR"), continued...

- Must discount the forecast of P&L to the present

Value at Risk ("VaR"), continued...

- Must discount the forecast of P&L to the present

$$\text{P\&L}_{t+h} = D_{t,h} P_{t+h} - P_t \tag{24}$$

Value at Risk ("VaR"), continued...

- Must discount the forecast of P&L to the present

$$\text{P\&L}_{t+h} = D_{t,h}P_{t+h} - P_t \tag{24}$$

- The Value at Risk of a portfolio — denoted by $\text{VaR}^{\$}_{\alpha,h}$ — expressed in dollar terms:

$$\Pr(D_{t,h}P_{t+h} - P_t < q^{\$}_{\alpha,h}) = \alpha \tag{25}$$

$$\text{VaR}^{\$}_{\alpha,h} = -q^{\$}_{\alpha,h}, \tag{26}$$

Value at Risk ("VaR"), continued...

- Must discount the forecast of P&L to the present

$$\text{P\&L}_{t+h} = D_{t,h} P_{t+h} - P_t \tag{24}$$

- The Value at Risk of a portfolio — denoted by $\text{VaR}^{\$}_{\alpha,h}$ — expressed in dollar terms:

$$\begin{aligned} \Pr(D_{t,h} P_{t+h} - P_t < q^{\$}_{\alpha,h}) &= \alpha \qquad (25) \\ \text{VaR}^{\$}_{\alpha,h} &= -q^{\$}_{\alpha,h}, \qquad (26) \end{aligned}$$

where $q^{\$}_{\alpha,h}$ is the quantile of the discounted (theoretical, or unrealized) P&L distribution at time $t + h$.

Value at Risk ("VaR"), continued. . .

Value at Risk (“VaR”), continued. . .

- When expressed in terms of discounted returns

$$X_{t+h} = \frac{D_{t,h} P_{t+h} - P_t}{P_t}, \tag{27}$$

Value at Risk ("VaR"), continued...

- When expressed in terms of discounted returns

$$X_{t+h} = \frac{D_{t,h}P_{t+h} - P_t}{P_t}, \tag{27}$$

then

$$\Pr(X_{t+h} < q_{\alpha,h}) = \alpha \tag{28}$$

Value at Risk ("VaR"), continued...

- When expressed in terms of discounted returns

$$X_{t+h} = \frac{D_{t,h}P_{t+h} - P_t}{P_t}, \tag{27}$$

then

$$\Pr(X_{t+h} < q_{\alpha,h}) = \alpha \tag{28}$$

and

$$\mathrm{VaR}_{\alpha,h} = \begin{cases} -q_{\alpha,h} & \text{as a \% of portfolio value } P_t, \\ -q_{\alpha,h}P_t = \mathrm{VaR}^{\$}_{\alpha,h} & \text{in dollar terms,} \end{cases} \tag{29}$$

where $q_{\alpha,h}$ is the quantile of the discounted (theoretical, or realized) returns distribution at time $t + h$.

Normal Linear VaR

Normal Linear VaR

- Suppose that
$$X_{t+h} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu_h, \sigma_h^2), \tag{30}$$
where μ_h is the expected return and σ_h^2 is the returns variance for holding period h.

Normal Linear VaR

- Suppose that
$$X_{t+h} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu_h, \sigma_h^2), \tag{30}$$
where μ_h is the expected return and σ_h^2 is the returns variance for holding period h.
- We can re-write Equation (28):

$$\begin{aligned}
\Pr(X_{t+h} < q_{\alpha,h}) &= \Pr\left(\frac{X_{t+h} - \mu_h}{\sigma_h} < \frac{q_{\alpha,h} - \mu_h}{\sigma_h}\right) && (31)\\
&= \Pr\left(Z < \frac{q_{\alpha,h} - \mu_h}{\sigma_h}\right) && (32)\\
&= \Pr\left(Z < \Phi^{-1}(\alpha)\right) && (33)\\
&= \alpha, && (34)
\end{aligned}$$

where $Z \sim \mathcal{N}(0, 1)$, and Φ^{-1} is the inverse normal distribution function (*e.g.* $\Phi^{-1}(0.01) \approx 2.33$).

Normal Linear VaR, continued. . .

Normal Linear VaR, continued...

- From Equation (32) and (33) we have

$$\frac{q_{\alpha,h} - \mu_h}{\sigma_h} = \Phi^{-1}(\alpha), \tag{35}$$

Normal Linear VaR, continued...

- From Equation (32) and (33) we have

$$\frac{q_{\alpha,h} - \mu_h}{\sigma_h} = \Phi^{-1}(\alpha), \tag{35}$$

- and from Equation (29) we have $\mathrm{VaR}_{\alpha,h} = -q_{\alpha,h}$ and $\Phi^{-1}(\alpha) = \Phi^{-1}(1-\alpha)$ by the symmetry of the normal distribution,

Normal Linear VaR, continued...

- From Equation (32) and (33) we have

$$\frac{q_{\alpha,h} - \mu_h}{\sigma_h} = \Phi^{-1}(\alpha), \tag{35}$$

- and from Equation (29) we have $\mathrm{VaR}_{\alpha,h} = -q_{\alpha,h}$ and $\Phi^{-1}(\alpha) = \Phi^{-1}(1-\alpha)$ by the symmetry of the normal distribution,
- so

$$\mathrm{VaR}_{\alpha,h} = -\mu_h + \Phi^{-1}(1-\alpha)\sigma_h. \tag{36}$$

Normal Linear VaR, continued...

- From Equation (32) and (33) we have

$$\frac{q_{\alpha,h} - \mu_h}{\sigma_h} = \Phi^{-1}(\alpha), \tag{35}$$

- and from Equation (29) we have $\mathrm{VaR}_{\alpha,h} = -q_{\alpha,h}$ and $\Phi^{-1}(\alpha) = \Phi^{-1}(1-\alpha)$ by the symmetry of the normal distribution,
- so

$$\mathrm{VaR}_{\alpha,h} = -\mu_h + \Phi^{-1}(1-\alpha)\sigma_h. \tag{36}$$

- Equation (36) is therefore an analytic expression for the VaR under the assumption of Equation (30), and it is this analytic expression that makes the normal linear model so tractable.

How can we predict VaR?

How can we predict VaR?

- Consider a simple FX forward outright:

How can we predict VaR?

- Consider a simple FX forward outright:
- We seek the valuation & risk measures of in USD,

How can we predict VaR?

- Consider a simple FX forward outright:
- We seek the valuation & risk measures of in USD, *i.e.* where $S_t, F_{t,T-t}, f_{t,T-t}, K$ quoted as CCYUSD

How can we predict VaR?

- Consider a simple FX forward outright:
- We seek the valuation & risk measures of in USD, *i.e.* where $S_t, F_{t,T-t}, f_{t,T-t}, K$ quoted as CCYUSD
-

$$
\begin{aligned}
V^{\$}_{t,T-t} \Leftrightarrow V^{\$}(K, S_t, r_d, r_f, T-t) &= \pm N^{\mathrm{c}} \times e^{-r_d(T-t)} \times (\mathbb{E}_t[S_T] - K) \\
&= \pm N^{\mathrm{c}} \times e^{-r_d(T-t)} \times (F_{t,T-t} - K) \\
&= \pm N^{\mathrm{c}} \times e^{-r_d(T-t)} \times (S_t + f_{t,T-t} - K) \\
&= \pm N^{\mathrm{c}} \times e^{-r_d(T-t)} \times (S_t e^{(r_d - r_f)(T-t)} - K) \\
&= \pm N^{\mathrm{c}} \times \left(S_t e^{-r_f(T-t)} - K e^{-r_d(T-t)}\right)
\end{aligned}
$$

The Humble FX Forward Contract

The Humble FX Forward Contract

- Risk measures:

The Humble FX Forward Contract

- Risk measures:

$$\Delta V^{\$}_{t,\textsc{Spot}} \quad = \quad \frac{\partial V^{\$}_t}{\partial S_t} \times \Delta S$$

The Humble FX Forward Contract

- Risk measures:

$$\Delta V^{\$}_{t,\textsc{Spot}} \quad = \quad \frac{\partial V^{\$}_t}{\partial S_t} \times \Delta S = \pm N^{\mathrm{C}} \times e^{-r_f(T-t)} \times \Delta S$$

The Humble FX Forward Contract

- Risk measures:

$$\begin{aligned}
\Delta V^{\$}_{t,\textsc{Spot}} &= \frac{\partial V^{\$}_t}{\partial S_t} \times \Delta S = \pm N^{\mathrm{C}} \times e^{-r_f(T-t)} \times \Delta S \\
\Delta V^{\$}_{t,r_d} &= \frac{\partial V^{\$}_t}{\partial r_d} \times \Delta r_d
\end{aligned}$$

The Humble FX Forward Contract

- Risk measures:

$$\Delta V^{\$}_{t,\text{SPOT}} = \frac{\partial V^{\$}_t}{\partial S_t} \times \Delta S = \pm N^{\text{C}} \times e^{-r_f(T-t)} \times \Delta S$$

$$\Delta V^{\$}_{t,r_d} = \frac{\partial V^{\$}_t}{\partial r_d} \times \Delta r_d = \pm N^{\text{C}} \times K(T-t) \times \Delta r_d$$

The Humble FX Forward Contract

- Risk measures:

$$\begin{aligned}
\Delta V^{\$}_{t,\textsc{Spot}} &= \frac{\partial V^{\$}_t}{\partial S_t} \times \Delta S = \pm N^{\mathrm{C}} \times e^{-r_f(T-t)} \times \Delta S \\
\Delta V^{\$}_{t,r_d} &= \frac{\partial V^{\$}_t}{\partial r_d} \times \Delta r_d = \pm N^{\mathrm{C}} \times K(T-t) \times \Delta r_d \\
\Delta V^{\$}_{t,r_f} &= \frac{\partial V^{\$}_t}{\partial r_f} \times \Delta r_f
\end{aligned}$$

The Humble FX Forward Contract

- Risk measures:

$$\begin{aligned}
\Delta V^{\$}_{t,\textsc{Spot}} &= \frac{\partial V^{\$}_t}{\partial S_t} \times \Delta S = \pm N^{\mathrm{C}} \times e^{-r_f(T-t)} \times \Delta S \\
\Delta V^{\$}_{t,r_d} &= \frac{\partial V^{\$}_t}{\partial r_d} \times \Delta r_d = \pm N^{\mathrm{C}} \times K(T-t) \times \Delta r_d \\
\Delta V^{\$}_{t,r_f} &= \frac{\partial V^{\$}_t}{\partial r_f} \times \Delta r_f = \pm N^{\mathrm{C}} \times -S_t(T-t) \times \Delta r_f
\end{aligned}$$

The Humble FX Forward Contract

- Risk measures:

$$
\begin{aligned}
\Delta V^{\$}_{t,\textsc{Spot}} &= \frac{\partial V^{\$}_t}{\partial S_t} \times \Delta S = \pm N^{\mathrm{C}} \times e^{-r_f(T-t)} \times \Delta S \\
\Delta V^{\$}_{t,r_d} &= \frac{\partial V^{\$}_t}{\partial r_d} \times \Delta r_d = \pm N^{\mathrm{C}} \times K(T-t) \times \Delta r_d \\
\Delta V^{\$}_{t,r_f} &= \frac{\partial V^{\$}_t}{\partial r_f} \times \Delta r_f = \pm N^{\mathrm{C}} \times -S_t(T-t) \times \Delta r_f \\
\Delta V^{\$}_{t,\tau} &= \frac{\partial V^{\$}_t}{\partial \tau} \times \Delta \tau
\end{aligned}
$$

The Humble FX Forward Contract

- Risk measures:

$$\begin{aligned}
\Delta V^{\$}_{t,\textsc{Spot}} &= \frac{\partial V^{\$}_t}{\partial S_t} \times \Delta S = \pm N^{\mathrm{C}} \times e^{-r_f(T-t)} \times \Delta S \\
\Delta V^{\$}_{t,r_d} &= \frac{\partial V^{\$}_t}{\partial r_d} \times \Delta r_d = \pm N^{\mathrm{C}} \times K(T-t) \times \Delta r_d \\
\Delta V^{\$}_{t,r_f} &= \frac{\partial V^{\$}_t}{\partial r_f} \times \Delta r_f = \pm N^{\mathrm{C}} \times -S_t(T-t) \times \Delta r_f \\
\Delta V^{\$}_{t,\tau} &= \frac{\partial V^{\$}_t}{\partial \tau} \times \Delta \tau = \pm N^{\mathrm{C}} \times \left(-r_f S_t + r_d K\right) \times \Delta \tau
\end{aligned}$$

The Humble FX Forward Contract

- Risk measures:

$$
\begin{aligned}
\Delta V^{\$}_{t,\textsc{Spot}} &= \frac{\partial V^{\$}_t}{\partial S_t} \times \Delta S = \pm N^{\mathrm{C}} \times e^{-r_f(T-t)} \times \Delta S \\
\Delta V^{\$}_{t,r_d} &= \frac{\partial V^{\$}_t}{\partial r_d} \times \Delta r_d = \pm N^{\mathrm{C}} \times K(T-t) \times \Delta r_d \\
\Delta V^{\$}_{t,r_f} &= \frac{\partial V^{\$}_t}{\partial r_f} \times \Delta r_f = \pm N^{\mathrm{C}} \times -S_t(T-t) \times \Delta r_f \\
\Delta V^{\$}_{t,\tau} &= \frac{\partial V^{\$}_t}{\partial \tau} \times \Delta \tau = \pm N^{\mathrm{C}} \times \left(-r_f S_t + r_d K\right) \times \Delta \tau
\end{aligned}
$$

which, in vector form, becomes

The Humble FX Forward Contract

- Risk measures:

$$\begin{aligned}
\Delta V^{\$}_{t,\textsc{Spot}} &= \frac{\partial V^{\$}_t}{\partial S_t} \times \Delta S = \pm N^{\mathrm{C}} \times e^{-r_f(T-t)} \times \Delta S \\
\Delta V^{\$}_{t,r_d} &= \frac{\partial V^{\$}_t}{\partial r_d} \times \Delta r_d = \pm N^{\mathrm{C}} \times K(T-t) \times \Delta r_d \\
\Delta V^{\$}_{t,r_f} &= \frac{\partial V^{\$}_t}{\partial r_f} \times \Delta r_f = \pm N^{\mathrm{C}} \times -S_t(T-t) \times \Delta r_f \\
\Delta V^{\$}_{t,\tau} &= \frac{\partial V^{\$}_t}{\partial \tau} \times \Delta \tau = \pm N^{\mathrm{C}} \times \Big(-r_f S_t + r_d K\Big) \times \Delta \tau
\end{aligned}$$

which, in vector form, becomes

$$\Delta V^{\$}_t = \underbrace{\begin{bmatrix} \frac{\partial V^{\$}_t}{\partial S_t} \\ \frac{\partial V^{\$}_t}{\partial r_d} \\ \frac{\partial V^{\$}_t}{\partial r_f} \\ \frac{\partial V^{\$}_t}{\partial \tau} \end{bmatrix}}_{\nabla V^{\$}_t \Leftrightarrow \text{“Greeks”}} \odot \underbrace{\begin{bmatrix} \Delta S \\ \Delta r_d \\ \Delta r_f \\ \Delta \tau \end{bmatrix}}_{\varepsilon}$$

The Humble FX Forward Contract, continued. . .

The Humble FX Forward Contract, continued...

$$\Delta V_t^{\$} = \underbrace{\begin{bmatrix} \frac{\partial V_t^{\$}}{\partial S_t} \\ \frac{\partial V_t^{\$}}{\partial r_d} \\ \frac{\partial V_t^{\$}}{\partial r_f} \\ \frac{\partial V_t^{\$}}{\partial \tau} \end{bmatrix}}_{\nabla V_t^{\$} \Leftrightarrow \text{“Greeks”}} \odot \underbrace{\begin{bmatrix} \Delta S \\ \Delta r_d \\ \Delta r_f \\ \Delta \tau \end{bmatrix}}_{\varepsilon}$$

The Humble FX Forward Contract, continued...

$$\Delta V_t^{\$} = \underbrace{\begin{bmatrix} \frac{\partial V_t^{\$}}{\partial S_t} \\ \frac{\partial V_t^{\$}}{\partial r_d} \\ \frac{\partial V_t^{\$}}{\partial r_f} \\ \frac{\partial V_t^{\$}}{\partial \tau} \end{bmatrix}}_{\nabla V_t^{\$} \Leftrightarrow \text{“Greeks”}} \odot \underbrace{\begin{bmatrix} \Delta S \\ \Delta r_d \\ \Delta r_f \\ \Delta \tau \end{bmatrix}}_{\varepsilon} = \pm N^{C} \times \underbrace{\begin{bmatrix} e^{-r_f(T-t)} \\ K(T-t) \\ -S_t(T-t) \\ -r_f S_t + r_d K \end{bmatrix}}_{\nabla V_t^{\$} \Leftrightarrow \text{“Greeks”}} \odot \underbrace{\begin{bmatrix} \Delta S \\ \Delta r_d \\ \Delta r_f \\ \Delta \tau \end{bmatrix}}_{\varepsilon}$$

The Humble FX Forward Contract, continued...

$$\Delta V_t^{\$} = \underbrace{\begin{bmatrix} \frac{\partial V_t^{\$}}{\partial S_t} \\ \frac{\partial V_t^{\$}}{\partial r_d} \\ \frac{\partial V_t^{\$}}{\partial r_f} \\ \frac{\partial V_t^{\$}}{\partial \tau} \end{bmatrix}}_{\nabla V_t^{\$} \Leftrightarrow \text{“Greeks”}} \odot \underbrace{\begin{bmatrix} \Delta S \\ \Delta r_d \\ \Delta r_f \\ \Delta \tau \end{bmatrix}}_{\varepsilon} = \underbrace{\pm N^C \times \begin{bmatrix} e^{-r_f(T-t)} \\ K(T-t) \\ -S_t(T-t) \\ -r_f S_t + r_d K \end{bmatrix}}_{\nabla V_t^{\$} \Leftrightarrow \text{“Greeks”}} \odot \underbrace{\begin{bmatrix} \Delta S \\ \Delta r_d \\ \Delta r_f \\ \Delta \tau \end{bmatrix}}_{\varepsilon}$$

$$= \nabla V_t^{\$} \odot \varepsilon$$

The Humble FX Forward Contract, continued...

$$\Delta V_t^{\$} = \underbrace{\begin{bmatrix} \frac{\partial V_t^{\$}}{\partial S_t} \\ \frac{\partial V_t^{\$}}{\partial r_d} \\ \frac{\partial V_t^{\$}}{\partial r_f} \\ \frac{\partial V_t^{\$}}{\partial \tau} \end{bmatrix}}_{\nabla V_t^{\$} \Leftrightarrow \text{“Greeks”}} \odot \underbrace{\begin{bmatrix} \Delta S \\ \Delta r_d \\ \Delta r_f \\ \Delta \tau \end{bmatrix}}_{\varepsilon} = \underbrace{\pm N^C \times \begin{bmatrix} e^{-r_f(T-t)} \\ K(T-t) \\ -S_t(T-t) \\ -r_f S_t + r_d K \end{bmatrix}}_{\nabla V_t^{\$} \Leftrightarrow \text{“Greeks”}} \odot \underbrace{\begin{bmatrix} \Delta S \\ \Delta r_d \\ \Delta r_f \\ \Delta \tau \end{bmatrix}}_{\varepsilon}$$

$$= \nabla V_t^{\$} \odot \varepsilon$$

What is the aggregate effect of a perturbation to the value of the contract?

The Humble FX Forward Contract, continued...

$$
\begin{aligned}
\Delta V_t^{\$} &= \underbrace{\begin{bmatrix} \frac{\partial V_t^{\$}}{\partial S_t} \\ \frac{\partial V_t^{\$}}{\partial r_d} \\ \frac{\partial V_t^{\$}}{\partial r_f} \\ \frac{\partial V_t^{\$}}{\partial \tau} \end{bmatrix}}_{\nabla V_t^{\$} \Leftrightarrow \text{“Greeks”}} \odot \underbrace{\begin{bmatrix} \Delta S \\ \Delta r_d \\ \Delta r_f \\ \Delta \tau \end{bmatrix}}_{\varepsilon} = \underbrace{\pm N^C \times \begin{bmatrix} e^{-r_f(T-t)} \\ K(T-t) \\ -S_t(T-t) \\ -r_f S_t + r_d K \end{bmatrix}}_{\nabla V_t^{\$} \Leftrightarrow \text{“Greeks”}} \odot \underbrace{\begin{bmatrix} \Delta S \\ \Delta r_d \\ \Delta r_f \\ \Delta \tau \end{bmatrix}}_{\varepsilon} \\
&= \nabla V_t^{\$} \odot \varepsilon
\end{aligned}
$$

What is the aggregate effect of a perturbation to the value of the contract?

$$
V_t^{\$}(K, S_t + \Delta S, r_d + \Delta r_d, r_f + \Delta r_f, T - t + \Delta\tau) \approx V_t^{\$}(K, S_t, r_d, r_f, T - t) + (\nabla V_t^{\$})'\varepsilon
$$

The Humble FX Forward Contract, continued...

$$
\begin{aligned}
\Delta V_t^{\$} &= \underbrace{\begin{bmatrix} \frac{\partial V_t^{\$}}{\partial S_t} \\ \frac{\partial V_t^{\$}}{\partial r_d} \\ \frac{\partial V_t^{\$}}{\partial r_f} \\ \frac{\partial V_t^{\$}}{\partial \tau} \end{bmatrix}}_{\nabla V_t^{\$} \Leftrightarrow \text{“Greeks”}} \odot \underbrace{\begin{bmatrix} \Delta S \\ \Delta r_d \\ \Delta r_f \\ \Delta \tau \end{bmatrix}}_{\varepsilon} = \underbrace{\pm N^C \times \begin{bmatrix} e^{-r_f(T-t)} \\ K(T-t) \\ -S_t(T-t) \\ -r_f S_t + r_d K \end{bmatrix}}_{\nabla V_t^{\$} \Leftrightarrow \text{“Greeks”}} \odot \underbrace{\begin{bmatrix} \Delta S \\ \Delta r_d \\ \Delta r_f \\ \Delta \tau \end{bmatrix}}_{\varepsilon} \\
&= \nabla V_t^{\$} \odot \varepsilon
\end{aligned}
$$

What is the aggregate effect of a perturbation to the value of the contract?

$$
\underbrace{V_t^{\$}(K, S_t + \Delta S, r_d + \Delta r_d, r_f + \Delta r_f, T - t + \Delta \tau)}_{\text{Level (hypothesized from unrealized)}} \approx \underbrace{V_t^{\$}(K, S_t, r_d, r_f, T - t)}_{\text{Level (unrealized)}} + (\nabla V_t^{\$})' \varepsilon
$$

The Humble FX Forward Contract, continued...

$$
\begin{aligned}
\Delta V_t^{\$} &= \underbrace{\begin{bmatrix} \frac{\partial V_t^{\$}}{\partial S_t} \\ \frac{\partial V_t^{\$}}{\partial r_d} \\ \frac{\partial V_t^{\$}}{\partial r_f} \\ \frac{\partial V_t^{\$}}{\partial \tau} \end{bmatrix}}_{\nabla V_t^{\$} \Leftrightarrow \text{“Greeks”}} \odot \underbrace{\begin{bmatrix} \Delta S \\ \Delta r_d \\ \Delta r_f \\ \Delta \tau \end{bmatrix}}_{\varepsilon} = \pm N^C \times \underbrace{\begin{bmatrix} e^{-r_f(T-t)} \\ K(T-t) \\ -S_t(T-t) \\ -r_f S_t + r_d K \end{bmatrix}}_{\nabla V_t^{\$} \Leftrightarrow \text{“Greeks”}} \odot \underbrace{\begin{bmatrix} \Delta S \\ \Delta r_d \\ \Delta r_f \\ \Delta \tau \end{bmatrix}}_{\varepsilon} \\
&= \nabla V_t^{\$} \odot \varepsilon
\end{aligned}
$$

What is the aggregate effect of a perturbation to the value of the contract?

$$
\underbrace{V_t^{\$}(K, S_t + \Delta S, r_d + \Delta r_d, r_f + \Delta r_f, T - t + \Delta \tau)}_{\text{Level (hypothesized from unrealized)}} \approx \underbrace{V_t^{\$}(K, S_t, r_d, r_f, T - t)}_{\text{Level (unrealized)}} + (\nabla V_t^{\$})' \varepsilon
$$

or

The Humble FX Forward Contract, continued...

$$
\begin{aligned}
\Delta V_t^{\$} &= \underbrace{\begin{bmatrix} \frac{\partial V_t^{\$}}{\partial S_t} \\ \frac{\partial V_t^{\$}}{\partial r_d} \\ \frac{\partial V_t^{\$}}{\partial r_f} \\ \frac{\partial V_t^{\$}}{\partial \tau} \end{bmatrix}}_{\nabla V_t^{\$} \Leftrightarrow \text{“Greeks”}} \odot \underbrace{\begin{bmatrix} \Delta S \\ \Delta r_d \\ \Delta r_f \\ \Delta \tau \end{bmatrix}}_{\varepsilon} = \underbrace{\pm N^{C} \times \begin{bmatrix} e^{-r_f(T-t)} \\ K(T-t) \\ -S_t(T-t) \\ -r_f S_t + r_d K \end{bmatrix}}_{\nabla V_t^{\$} \Leftrightarrow \text{“Greeks”}} \odot \underbrace{\begin{bmatrix} \Delta S \\ \Delta r_d \\ \Delta r_f \\ \Delta \tau \end{bmatrix}}_{\varepsilon} \\
&= \nabla V_t^{\$} \odot \varepsilon
\end{aligned}
$$

What is the aggregate effect of a perturbation to the value of the contract?

$$
\underbrace{V_t^{\$}(K, S_t + \Delta S, r_d + \Delta r_d, r_f + \Delta r_f, T - t + \Delta \tau)}_{\text{Level (hypothesized from unrealized)}} \approx \underbrace{V_t^{\$}(K, S_t, r_d, r_f, T - t)}_{\text{Level (unrealized)}} + (\nabla V_t^{\$})' \varepsilon
$$

or

$$
\underbrace{V_t^{\$}(K, S_t + \Delta S, r_d + \Delta r_d, r_f + \Delta r_f, T - t + \Delta \tau) - V_t^{\$}(K, S_t, r_d, r_f, T - t)}_{\text{Hypothesized P\&L — the first-order difference — is stationary}} \approx (\nabla V_t^{\$})' \varepsilon
$$