The Stochastic Discount Factor and Generalized Method of Moments

Empirical Asset Pricing

Mads Markvart Kjær

Department of Economics and Business Economics, Aarhus University, CREATES

E-mail: mads.markvart@econ.au.dk

Spring 2022

Recap

- **Last lecture**, Jonas talked about **time-series predictability**
- **Today**, we will focus on the building block (or theory) of **cross-sectional predictability**, i.e., the **Stochastic Discount Factor** (SDF)

Why do we care?

Central asset pricing **questions**:

- Why do **different** assets give **different** (expected) **returns**?
- Is a **certain risk priced** in financial markets?
- How do we **interpret** compensation/**risk premia**?

➨ For trying to **answer** these questions, the dominant approach is to use the **stochastic discount factor**

- After today's lecture, we will all have a **new purpose in life**, i.e., **finding** the **stochastic discount factor**!

Basic properties

- A **SDF**, M_t, is a **random variable** with the following three **properties**:

  1. M_t has **finite variance**
  2. M_t is **strictly positive** (at least under no arbitrage)
  3. The **price of asset i** is given as

$$p_{i,t}(X_{i,t+1}) = \mathbb{E}_t(M_{t+1}X_{i,t+1}) \tag{1}$$

SDFs are everywhere!

Contents lists available at ScienceDirect

Journal of Empirical Finance

journal homepage: www.elsevier.com/locate/jempfin

Cross-sectional return dispersion and currency momentum☆

Jonas N. Eriksen

CREATES, Department of Economics and Business Economics, Aarhus University, Fuglesangs Allé 4, 8210 Aarhus V, Denmark

3.1. Methodology

In the absence of arbitrage, risk-adjusted currency excess returns have a price of zero and satisfy the basic Euler equation

$$E_t\left[M_{t+1} RX^j_{t+1}\right] = 0, \tag{3}$$

where RX^j_{t+1} is the excess return on currency portfolio j at time $t+1$ and M_{t+1} is a stochastic discount factor (SDF) that is linear in the risk factors f_{t+1}

$$M_{t+1} = 1 - b'\left(f_{t+1} - \mu_f\right), \tag{4}$$

where b is a vector of factor loadings and μ_f denotes factor means. This specification implies a beta pricing model

$$E\left[RX^j_{t+1}\right] = \lambda' \beta^j, \tag{5}$$

where the expected excess currency return depends on the factor risk prices λ and the corresponding factor betas β^j. The factor price

SDFs are everywhere!

The Journal of FINANCE

The Journal of THE AMERICAN FINANCE ASSOCIATION

THE JOURNAL OF FINANCE • VOL. LXVII, NO. 2 • APRIL 2012

Carry Trades and Global Foreign Exchange Volatility

LUKAS MENKHOFF, LUCIO SARNO, MAIK SCHMELING, and ANDREAS SCHRIMPF*

We denote excess returns of portfolio i in period $t+1$ by rx^i_{t+1}.[17] The usual no-arbitrage relation applies so that risk-adjusted currency excess returns have a price of zero and satisfy the basic Euler equation

$$\mathbb{E}\left[m_{t+1} rx^i_{t+1}\right] = 0, \tag{5}$$

with a linear SDF given by $m_t = 1 - b'(h_t - \mu)$, where h denotes a vector of risk factors, b is the vector of SDF parameters, and μ denotes the factor means. This specification implies a beta pricing model where expected excess returns depend on factor risk prices λ and risk quantities β_i, which are the regression betas of portfolio excess returns on the risk factors:

$$\mathbb{E}[rx^i] = \lambda' \beta_i, \tag{6}$$

Currency Premia and Global Imbalances

Pasquale Della Corte
Imperial College London and Centre for Economic Policy Research

Steven J. Riddiough
University of Melbourne

Lucio Sarno
City University London and Centre for Economic Policy Research

4.1 Methodology

We denote the discrete excess returns on portfolio j in period t as RX_t^j. In the absence of arbitrage opportunities, risk-adjusted excess returns have a price of zero and satisfy the following Euler equation:

$$E_t[M_{t+1} RX_{t+1}^j] = 0, \tag{4}$$

with a stochastic discount factor (SDF) linear in the pricing factors f_{t+1}, given by

$$M_{t+1} = 1 - b'(f_{t+1} - \mu), \tag{5}$$

where b is the vector of factor loadings, and μ denotes the factor means. This specification implies a beta pricing model in which the expected excess return on portfolio j is equal to the factor risk price λ times the risk quantities β^j. The beta pricing model is defined as

Outcome of lecture

After the lecture, you should have

- **knowlegde** and **understading** of
  - The **stochastic discount factor** (SDF), its **existence** and **use in asset pricing**, its **implications** for **arbitrage** and market **completeness**, and its **relation** to the **investor's marginal utility**
- and be able to
  - **Discuss** and **estimate** the SDF using **common empirical methods**

➜ The questions you can (try to) answer using the SDF is not only exciting, but **the theory** is also **highly relevant** for the **exam**!

- First, we do need to distinguish between **complete** and **incomplete** markets

Complete markets

Complete markets, cont.

To explain the idea of **complete** markets, consider the **following setup**:

- **No transactions costs** and **perfect information** (no frictions)
- A **discrete-state** model with S many **states of the world**, $s = 1, \ldots, S$, each with **probability** $\pi(s)$
- For each **state** s, there exists a **contingent claim** that pays \$**1** in state s and **nothing** in any other state (This is also known as an Arrow-Debreu asset)
- The **price** of the asset is $q(s)$

Complete markets

Properties of complete markets

- **All possible bets** of the **future states of the world** can be constructed using the contingent claims
- **Prices** on all contingent claims are **strictly positive**, $q(s) > 0$

- If $q(s) \leq 0$, we have an arbitrage oppotunity
  - Suppose $q(s) \leq 0$. The investors “buys" the asset for either **nothing** or even receives a **positive** payoff today and gets an asset that has
    1. **non-zero probability** for receiving a **positive payoff** if state s realizes in the next period,
    2. **zero probability** for a **negative payoff** in any future state

$\Rightarrow$ **infinitely attractive** investment

The fundamental equation of asset pricing

- The assets are only **distinguished** by their **state-dependent** payoffs $X(s)$, $s = 1, \ldots, S$
- Given the **finite state-space**, **all** assets can be **replicated** using **bundles of contingent claims**
- Under **no-arbitrage**, the **price of an asset** with payoff X is given as

$$p_i(X) = \sum_{s=1}^{S} q(s) X_i(s). \tag{2}$$

- Also known as **Cochrane's happy meal theorem**

The fundamental equation of asset pricing

Law of one price (intuitively)

The **law of one price** says, intuitively, that **two assets** with **identical** payoffs (characteristics) in every state **must have the same price**

- If this **does not hold**, it would imply **arbitrage opportunities**
- Why? Suppose the contrary, that is,

$$p_1 > p_2, \tag{3}$$

  but identical across all states.
- Buy asset 2, sell asset 1 yields:
  - $p_1 - p_2 > 0$ today
  - zero in next period with probability 1

The fundamental equation of asset pricing

- So a **violation of the law of one price** leads to **arbitrage**
- **Arbitrage** does, however, **not** necessarily lead to a **violation of the law of one price**

The fundamental equation of asset pricing

- To get an **expectational expression**, multiply (2) by $1 = \pi(s)/\pi(s)$

$$p(X) = \sum_{s=1}^{S} \pi(s) \frac{q(s)}{\pi(s)} X(s) = \sum_{s=1}^{S} \pi(s) M(s) X(s), \tag{4}$$

where $M(s) = q(s)/\pi(s)$ is **defined** as the **SDF**

Relation to risk-neutral probabilities

- To see the **connection** between the **SDF** and **risk-neutral probability**, multiply Eq. (4) by $1 = R_f / R_f$

$$p(X) = \sum_{s=1}^{S} \pi(s) M(s) R_f \frac{X(s)}{R_f} = \sum_{s=1}^{S} \pi^*(s) \frac{X(s)}{R_f}, \tag{5}$$

  where $\pi^*(s) = \pi(s) M(s) R_f$ is the **risk-neutral probability**
- **One-to-one mapping** between **SDF** and **risk-neutral probabilities**!

Back to the SDF

$$p(X) = \sum_{s=1}^{S} \pi(s) M(s) X(s), \tag{6}$$

where $M(s) = q(s)/\pi(s)$ is defined as the SDF

The fundamental equation of asset pricing

The **fundamental equation of asset pricing** reads

$$p(X) = \mathbb{E}[MX]. \tag{7}$$

Since $q(s), \pi(s) > 0$, it follows that $M(s) > 0$

- Let us **explore/recap** an application of it ➨ **consumption-based asset pricing (CCAPM)**

The representative agent

- A **market equilibrium** consists of **many** (heterogeneous) **investors**, each optimizing their utility
- Wouldn't it be **nice** if we could **simplify the market** into a **single representative agent** and get the **same equilibrium**?

The aggregation property of the economy

- If **markets are complete**, financial markets have the **aggregation property**
- That is, **equilibrium prices** are the same as in a **hypothetical representative-agent economy**
- …, and we can **work** with a **single representative agent**

- Consumption-based asset pricing models frequently aggregate individual investors into a single utility-maximizing (representative) agent whose **utility** derives from **aggregate (per capita) consumption**

The representative agent, a small note

- There exists **several formulations** of the **utility-maximizing intertemporal choice problem** of the representative investor
- **Not all** formulations **lead** to this **aggregation property**, yet we will not deal with this further, see Campbell (2017) p. 89

Consumption-based asset pricing framework

- Let $u(c_t)$ be the **concave**, **time-separable utility function**, where c denotes aggregate consumption (per capita)
- Each period, the investor **chooses** between **consumption** and **investing** (for future consumption) to optimally smooth consumption

The maximization problem

The **representative agent maximizes**

$$\max \sum_{t=1}^{T} \delta^t \mathbb{E}[u(c_t)|\mathcal{F}_t], \tag{8}$$

subject to budget constraints, which we leave unspecified for now, and a large T

- $\delta = (1+\tau)^{-1}$ is a (deterministic) **subjective discount factor**, and τ is the **subjective time preference rate**. The smaller δ, the more **impatient** the investor is ➨ it prefers consumption **now** versus in the **future**
- $\mathcal{F}_t$ is the time-t **filtration** (**information set** available to the investor).

Consumption-based asset pricing framework

Euler equation (version 1)

The **solution** to the **maximization problem** is

$$u'(c_t)P_{i,t} = \mathbb{E}\left[\delta u'(c_{t+1})(P_{i,t+1} + D_{i,t+1})|\mathcal{F}_t\right]. \tag{9}$$

where $P_{i,t}, D_{i,t}$ is the **price** and **dividend** paid by **asset** i at time t and u' is the first **derivative** of the **utility function** w.r.tc (marginal utility)

- The Euler equation is optimum for the representative investor's consumption and portfolio choice problem
- It **equates** marginal **cost** and **benefit** of current versus future consumption:
  1. LHS: **Marginal utility loss** in period t from buying one additional unit of the asset instead of consuming today
  2. RHS: **Expected discounted marginal utility gain** associated with buying an additional unit of the asset instead of consuming today

Consumption-based asset pricing framework

Euler equation (version 2)

The **solution** to the maximization problem can be **rewritten** to

$$P_{it} = \mathbb{E}\left[\delta \frac{u'(c_{t+1})}{u'(c_t)} (P_{i,t+1} + D_{i,t+1}) | \mathcal{F}_t \right], \tag{10}$$

or, **equivalently**,

$$1 = \mathbb{E}\left[\delta \frac{u'(c_{t+1})}{u'(c_t)} R_{i,t+1} | \mathcal{F}_t \right], \tag{11}$$

where $R_{i,t+1} = (P_{i,t+1} + D_{i,t+1}) / P_{i,t}$ is the (gross) return on asset i.

- Since we consider a period-by-period optimization problem, the payoff of asset i is $X_{i,t+1} = P_{i,t+1} + D_{i,t+1}$

Consumption-based asset pricing framework

- This **matches** the structure of the **simple fundamental equation** of asset pricing in (7)

Theorem: SDF in consumption-bassed asset pricing

In a **discrete-time**, **complete** market economy with a single consumption good, let δ be the time subjective discount factor and u the utility function of the representative individual with time-separable utility. Then (the process)

$$M_{t+1} = \delta \frac{u'(c_{t+1})}{u'(c_t)}, \quad t = 0,1,\ldots,T \tag{12}$$

is an **SDF** (at all time points).

Consumption-based asset pricing framework

- As such, the **price** of asset i at any time-point is the discounted value of the future payoff
- The discounting is the **marginal rate of substitution** between time t and $t+1$ consumption ➨ the growth in marginal utility.
- with a **large growth** in marginal utility, any future payoff is **highly valued** and the **price today** (expected return) will be **higher** (lower), and vice versa
- The **functional** form of M_{t+1} **depends** on the **choice of the utility function** and is extremely scrutinized in the academic literature
- ..., for now, we will consider the most famous (and the most simple) example, i.e., power utility

Consumption-based asset pricing framework

Power-utility Euler equation

Suppose the **representative investor** has time-separable power utility

$$u(c) = \frac{c^{1-\rho}}{1-\rho}, \tag{13}$$

where $\rho > 0$ is the relative risk aversion (coefficient). The Euler equation is then

$$P_{i,t} = \mathbb{E}\left[\delta \left(\frac{c_{t+1}}{c_t}\right)^{-\rho} (P_{i,t+1} + D_{i,t+1}) | \mathcal{F}_t\right]. \tag{14}$$

- Obtaining this **Euler equation** follows simply from computing the derivative of the utility function and inserting in the SDF of the consumption-based asset pricing framework in (12)

Consumption-based asset pricing framework

- A compact notation is, thus,

$$1 = \mathbb{E}_t[M_{t+1} R_{i,t+1}], \tag{15}$$

using (12), and where we use subscript t to indicate **conditional moments**.

Central consumption-based asset pricing equation

The **central consumption-based asset pricing equation** is

$$\mathbb{E}_t[r_{i,t+1}] - r_{f,t+1} = -(1 + r_{f,t+1})\text{Cov}_t[M_{t+1}, r_{i,t+1}], \tag{16}$$

where $r_{i,t+1}$ is the simple return, $r_{i,t+1} = R_{it+1} - 1$. Equivalently, using (12),

$$\mathbb{E}_t[r_{i,t+1}] - r_{f,t+1} = -\delta(1 + r_{f,t+1})\text{Cov}_t\left[\frac{u'(c_{t+1})}{u'(c_t)}, r_{i,t+1}\right]. \tag{17}$$

Consumption-based asset pricing framework

Consumption-based asset pricing logic

Assets with $\text{Cov}_t\left[\frac{u'(c_{t+1})}{u'(c_t)}, r_{i,t+1}\right] < 0$ earn **higher expected excess returns**

- Note that $\frac{u'(c_{t+1})}{u'(c_t)}$ is inversely related to the business cycle:
  1. **high** during **recessions** (when consumption is low)
  2. **low** during **expansions** (when consumption is high)
- If $\text{Cov}_t\left[\frac{u'(c_{t+1})}{u'(c_t)}, r_{i,t+1}\right] < 0$, asset i **pays off poorly in bad states** and well in good states, making it undesirable for consumption smoothing purpose
- If $\text{Cov}_t\left[\frac{u'(c_{t+1})}{u'(c_t)}, r_{i,t+1}\right] > 0$, asset i **provides consumption insurance by paying off in bad states** when the investor values additional consumption most highly

Incomplete markets

Incomplete markets

- What if markets are **incomplete**?
- Rather than **deriving a specific SDF** as in the consumption-based framework, we will now **work backward** (and be **more general**)
- Essentially, an **SDF** is just defined as the random variable that makes the following **representations true**

$$P_{i,t} = \mathbb{E}_t[M_{t+1}X_{i,t+1}] \quad \text{and} \quad 1 = \mathbb{E}_t[M_{t+1}R_{it+1}],\ \forall i,t$$

- When can we find such SDF, M_{t+1}?
- Can we **use** this representation **without implicitly assuming** all the structure of the investors, utility functions, complete markets, etc.?
- The **short answer** will be **yes**! …, under some conditions.

Incomplete markets

- **Suppose** we **observe** a set of asset payoffs X and prices P

Payoff space

The **payoff space**, denoted Ξ, is defined as the set of all the payoffs that investors can buy, including combining various assets

- To **obtain existence** of (at least one) **SDF**, we need to put some high-level structure on the economy

Incomplete markets

Assumptions

We make the following two assumptions:

1. **Portfolio formation**: $X_1, X_2 \in \Xi \Rightarrow X_p \equiv aX_1 + bX_2 \in \Xi$ for any real-valued a,b
2. **Law of one price**: $P(X_p) \equiv P(aX_1 + bX_2) = aP(X_1) + bP(X_2)$.

- Assumption 1 is quite restrictive in the sense that it rules out shorting constraints (by allowing $a,b < 0$), bid-ask spreads, leverage limitations, etc.
- …those can, however, be incorporated at the cost of complexity

Incomplete markets

Assumptions

We make the following two assumptions:

1. **Portfolio formation**: $X_1, X_2 \in \Xi \Rightarrow X_p \equiv aX_1 + bX_2 \in \Xi$ for any real-valued a,b
2. **Law of one price**: $P(X_p) \equiv P(aX_1 + bX_2) = aP(X_1) + bP(X_2)$.

- It states that it does not matter how one forms the payoff X_p - the price of a Happy Meal should be the sum of its constituents
- Assumption 2 is quite restrictive in the sense that it rules out the effect of packaging - a package is worth only what it contains and now how it is, e.g., branded

Existence of an SDF

Theorem: Existence of an SDF

Given portfolio formation (Assumption 1) and the law of one price (Assumption 2), there exists a payoff $X^* \in \Xi$ such that

$$P(X) = \mathbb{E}[X^* X], \quad \forall X \in \Xi. \tag{18}$$

- That is, under **Assumption 1 and 2**, X^* **satisfies the fundamental equation of asset pricing**, (7), **without the positivity property, $X^* > 0$, ensured**
- As such, X^* is an SDF ➨ in (in)complete markets it is (not) unique.
- It also goes the other way, i.e., the **existence of an SDF** implies **Assumption 1 and Assumption 2**

Positivity of the SDF

- While **Assumption 1 and 2** ensure the **existence** of an SDF, it does **not guarantee positivity**
- Why do we need it to be positive? It naturally results from any sort of **utility maximization**
- Recall,

$$M_{t+1} = \delta \frac{u'(c_{t+1})}{u'(c_t)}. \tag{19}$$

- Since $\delta > 0$ and $u'(c) > 0$ (unreasonable to think that people will get more utility from consuming less), $M_{t+1} > 0$
- But positivity of the SDF also rules out negative prices for assets that pay positive payoffs

Positivity of the SDF

Absence of arbitrage

A payoff space Ξ and pricing function $P(X)$ have absence of arbitrage if every payoff with $X \geq 0$ with certainty and if every payoff with positive $X > 0$ with some positive probability has positive price $P(X) > 0$.

- This definition is slightly different from the one given in Campbell (2017) but it is more intuitive
- It means that you cannot get a portfolio for free that *might* pay off positively, but will never certainly cost you anything.

Positivity of the SDF

Theorem: Positivity and existence of the SDF

1. $P = \mathbb{E}[MX]$ and $M(s) > 0 \Rightarrow$ absence of arbitrage.
2. Absence of arbitrage $\Rightarrow \exists M$ such that $P = \mathbb{E}[MX]$ and $M(s) > 0$.

- That is, a **positive SDF exists** if and only if **markets are free of arbitrage**. If so, all assets can be priced according to the fundamental equation of asset pricing in (7)

Why does all this theory matter for an empirical exercise?

- In the end, every choice we make most be due to some maximization exercise
- A natural way to motivate a risk factor is how it relates to the maximization problem of the representative investor:

$$\max \sum_{t=1}^{T} \delta^t \mathbb{E}[u(c_t)|\mathcal{F}_t], \tag{20}$$

- If a **risk factor** has an impact on **risk aversion**, **consumption (opportunities)**, **investment opportunities**, or the time-separable **discount function**, it **directly affects** the **SDF** and **expected returns**

Recap:

- A **postive SDF exists** if and only if **prices admits no arbitrage**
- Given no-arbitrage, the **SDF is unique** if and only if **markets are complete**

SDF and β representation

SDF-talk: Properties

- The **fundamental pricing equation**, using the SDF, is one type of **representation of asset pricing**
- **Two others exist**; β representation and mean-variance frontier representation

“All are equivalent" representation theorem

…both representations are equivalent

1. SDF $\Rightarrow \beta$
2. $\beta \Rightarrow$ SDF

- That is, if an **SDF exists**, we can always **find a β representation for asset returns**, and vice versa
- Additional details can be found in Cochrane (2009) Ch. 6

SDF-talk: SDF $\Rightarrow \beta$

- Recall the **fundamental asset pricing equation** in returns

$$1 = \mathbb{E}_t[M_{t+1}R_{i,t+1}] \tag{21}$$

  expressed in a discrete-time multi-period fashion

- Recall that if (21) holds for all t, it must also hold unconditionally (use law of iterated expectations), such that

$$1 = \mathbb{E}[M_{t+1}R_{i,t+1}]. \tag{22}$$

- We will discuss the **unconditional** implications of the **conditional** models further later
- In the following, we will work with this unconditional implication ➨ it essentially puts focus on average returns

SDF-talk: SDF $\Rightarrow \beta$

Risk-free rate in SDF form

Consider the **risk-free asset**, denote it by "$i = f$", with payoff $X_{f,t+1} = 1$ in all states, with **certainty**. By the fundamental asset pricing equation we must then have

$$\mathbb{E}[P_{f,t}] = P_{f,t} = \mathbb{E}[M_{t+1}], \tag{23}$$

such that

$$R_{ft+1} = \mathbb{E}[M_{t+1}]^{-1}. \tag{24}$$

SDF-talk: SDF $\Rightarrow \beta$

- It follows by **general covariance rules** that

$$\begin{aligned} 1 &= \mathbb{E}[M_{t+1}R_{i,t+1}] \\ &= \mathbb{E}[M_{t+1}]\mathbb{E}[R_{i,t+1}] + \mathrm{Cov}[M_{t+1}, R_{i,t+1}]. \end{aligned} \tag{25}$$

- …such that

$$\mathbb{E}[R_{i,t+1} - R_{f,t+1}] = -R_{f,t+1}\mathrm{Cov}[M_{t+1}, R_{i,t+1}]. \tag{26}$$

- The return on any asset is:
  - The risk-free return
  - A term that informs about the **co-variation** between the **SDF** and **returns** (This is where all the intuition in asset pricing models comes from!)

SDF-talk: SDF $\Rightarrow \beta$

SDF $\Rightarrow \beta$ representation

It follows from (26) that (multiply by $\frac{\text{Var}[M_{t+1}]}{\text{Var}[M_{t+1}]}$)

$$\mathbb{E}[R_{i,t+1} - R_{f,t+1}] = \beta_{i,M}\gamma_M, \tag{27}$$

where

$$\beta_{i,M} = \frac{\text{Cov}[M_{t+1}, R_{i,t+1}]}{\text{Var}[M_{t+1}]} \tag{28}$$

is the (single) **regression coefficient** of any asset return R_{it+1} on the SDF, and

$$\gamma_M = -R_{f,t+1}\text{Var}[M_{t+1}] \tag{29}$$

is the **factor risk premium**, noting that $R_{f,t+1}$ is known with certainty. (*do not confuse the subscript M with "market" ➨ it is due to the SDF denoted by M*)

SDF-talk: SDF $\Rightarrow \beta$

- This relates directly to the **intuition** presented in the consumption-based framework ➨ **expected excess returns** are **linear** in the regression βs of asset returns on $M_{t+1} = (c_{t+1}/c_t)^{-\rho}$
- Typically, γ_M is treated as a **free parameter** and **estimated** in empirical evaluations of factor models, however according to theory it should equal $-R_{ft+1}\text{Var}[(c_{t+1}/c_t)^{-\rho}] < 0$

β representation implications

- For a choice/model of SDF, a β representation is thus implied (and can be estimated)
- **Differences in expected excess returns** among a cross-section of assets must be explained by **differences in their βs (risks)**
- This **defines** the **empirical approaches** to **estimation of asset pricing models** which we will see/cover in the next lecture

SDF-talk: $\beta \Rightarrow$ SDF

- **Suppose** we have an **expected return model** in β representation (for instance, the CAPM). What SDF does this imply?

$\beta \Rightarrow$ SDF representation

A β representation of expected returns are equivalent to linear models for the SDF as per

$$M_{t+1} = a - b' f_{t+1}, \tag{30}$$

where a,b are parameters and f_{t+1} the risk factors. We use negative b for expositional reasons, see e.g. the example with CCAPM below in (34)

SDF-talk: $\beta \Rightarrow$ SDF

- Typically, we make **two convenient assumptions** that are **without loss of generality**:

Assumptions (w.l.o.g.)

1. **De-meaned factors**: We assume that factors are de-meaned such that $\mathbb{E}[f_{t+1}] = 0$. This implies that $\mathbb{E}[M_{t+1}] = \mathbb{E}[a - b' f_{t+1}] = a$.
2. **Normalization**: We normalize the mean of the SDF to unity, i.e. $\mathbb{E}[M_{t+1}] = 1$, which under Assumption 1 just above implies that $a = 1$

- Note, we are only able to **identify the SDF** up to the **scale of a constant** since $M_{t+1} = a(1 - (b/a)' f_{t+1})$

SDF-talk: $\beta \Rightarrow$ SDF

$\beta \Rightarrow$ SDF representation theorem

Suppose Assumptions 1 (de-meaned factors) and 2 (normalization) holds. Given the following β representation,

$$\mathbb{E}[R_{it+1} - R_{ft+1}] = \beta_i' \gamma, \tag{31}$$

where β are multiple regression coefficients of excess returns on the factors, we can always find b such that

$$M_{t+1} = 1 - b' f_{t+1} \tag{32}$$

with $\mathbb{E}[M_{t+1}(R_{it+1} - R_{ft+1})] = 0$.

- Also, given (32), we can always find a γ such that (31) holds

SDF-talk: Interpretation

- From **the β representation**, it is clear that γ may be interpreted as the **price of the factor risk**, or the **factor risk premium**
- For every unit β, the expected excess return increases by γ

The fundamental research question

As such, a test of $\gamma \neq 0$ is often called **a test** of whether **the factor is "priced" in the financial markets**. This is typically the main research question posed in studies in empirical asset pricing

SDF-talk: Interpretation via CCAPM

Example of CCAPM

- The CCAPM (approximately) stipulates that there exist a single risk factor, which is the logarithmic growth rate in aggregate consumption, denoted by $f_{t+1} = \tilde{c}_{t+1}$
- Suppose it is de-meaned and that Assumption 2 is invoked, such that

$$M_{t+1} = 1 - b\tilde{c}_{t+1}. \tag{33}$$

- It then follows from (26) that

$$\begin{aligned}
\mathbb{E}[R_{i,t+1} - R_{f,t+1}] &= -R_{f,t+1}\mathsf{Cov}[M_{t+1}, R_{i,t+1}] \\
&= -R_{f,t+1}\mathsf{Cov}[1 - b\tilde{c}_{t+1}, R_{i,t+1}] \\
&= R_{f,t+1} b \mathsf{Cov}[\tilde{c}_{t+1}, R_{i,t+1}] \\
&= \beta_i^c \gamma^c,
\end{aligned} \tag{34}$$

where $\beta_i^c = \mathsf{Cov}[\tilde{c}_{t+1}, R_{i,t+1}]/\mathsf{Var}[\tilde{c}_{t+1}]$ and $\gamma^c = R_{f,t+1} b \mathsf{Var}[\tilde{c}_{t+1}]$

SDF-talk: Interpretation via CCAPM

Example of CCAPM

- If $b > 0$ higher consumption growth reduces marginal utility growth. In this case, we have that $\gamma^c > 0$ (assuming $R_{ft+1} > 0$)
- That is, $\beta_i^c > 0$ is compensated/priced in the financial markets

SDF-talk: Uniformity and challenges

- So, **an asset pricing framework** that initially seemed to require a **lot of structure** (the representative, utility-maximising agent in the consumption-based framework) turns out to require **minimal structure**
- **Under few appropriate assumptions**, we can **always** start an analysis by writing $P_{it} = \mathbb{E}_t[M_{t+1}X_{t+1}]$ or $0 = \mathbb{E}_t[M_{t+1}(R_{it+1} - R_{ft+1})]$, **for any asset** (equity, bond, currency (we will later return to the asset class and how it relates to SDFs), house, cryptocurrency, you name it)

SDF-talk: Uniformity and challenges

- …and **this does not require** any **assumptions** on **market completeness**, **contingent-claim** or **representative agent**
- Of course, **this it not without a cost**: **all the economic, statistical, and predictive content** comes from **picking the SDF** model, i.e. $M_{t+1} = h(\text{data}_{t+1},\theta)$, for some function $h(\cdot,\theta)$

GMM

Generalized Methods of Moments (GMM)

- **Generalized Methods of Moments** (GMM) is an **estimation** principle, using **moment conditions** to enable identification.
- **Nests OLS**, **instrumental variables**, and **MLE**
- **Moment conditions** are of the form

$$\mathbb{E}[G(\text{data}_t, \theta)] = 0, \tag{35}$$

  where $G(\cdot)$ is a N-dimensional **function of data** and a K-dimensional vector of **parameters**, θ, that is to be estimated.
- See the relationship to (14), the CCAPM?

$$P_{it} = \mathbb{E}\left[\delta\left(\frac{c_{t+1}}{c_t}\right)^{-\rho}(P_{it+1} + D_{it+1})|\mathcal{F}_t\right].$$

Motivation

- **Robust** to assumptions about **homoskedasticity and autocorrelation**
- **Robust** to **distributional assumptions**
- **Can handle non-linear models**
- **Can handle economic models** that are formulated directly as **moment conditions**
- …and it is **extremely useful** for estimating (linear) β **represented factor models**, taking into account important statistical/empirical issues such as **errors-in-variables, autocorrelation, and heteroscedasticity**

Estimation approach

- To **estimate model parameters**, we consider the **sample average counterpart** of the moment conditions, called the **object function**:

$$g_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} G(\text{data}_t,\theta), \tag{36}$$

  where T is the sample time series dimension
- Note that $g_T(\theta) \xrightarrow{p} \mathbb{E}[G(\text{data}_t,\theta)]$ as $T \to \infty$.

Estimation principle (intuition)

GMM estimates parameters as those that make the **object function**, $g_T(\theta)$, **as close** to the ones implied by the **moment conditions**, i.e., 0.

Estimation approach

- We **need at least as many moment conditions** as we have **model parameters**, $N \geq K$:
    1. If $N < K$, the model is **not identified**
    2. If $N = K$, the model is **exactly identified**, sometimes with an analytical solutions if $G(\cdot)$ is linear in θ
    3. If $N > K$, the system is **overidentified** and **numerical optimization** is needed
- We need a way to **weight** each moment condition in the estimation, denoting this **weighting matrix** by A_T

Estimation approach

Estimation principle (formally)

For **a given choice** of **weighting matrix** (discussed below), GMM estimates parameters by **minimizing** a quadratic form of the weighted sample moment condition as per

$$\hat{\theta} = argmin_{\theta} g_T(\theta)' A_T g_T(\theta) \tag{37}$$

- For **any choice of weighting matrix** (e.g. the identity matrix), the **GMM estimator is consistent**, $\hat{\theta} \xrightarrow{p} \theta$ as $T \to \infty$
- The estimation procedure is often done in two or several steps, coined **two-stage** and **iterated** GMM
- To understand why, we need to understand the choice of weighting matrix

Choice of weighting matrix

- In the case of exact identification, we have that all moment conditions can be set equal to zero
- In the overidentified case, this is no longer possible
- The **weighting matrix** **determines the weight** each moment should have when estimating the parameters ➨ a very **important choice**

Symmetric (or equal) weights

If the weighting matrix is set to the identity matrix, it puts equal emphasis on all moment conditions, that is,

$$A_T = I_N,$$

where I_N is the $N \times N$-dimensional identity matrix

Choice of weighting matrix

- One particular choice of A_T is **optimal** in a **statistical sense**
- ...in the sense that the resulting GMM estimator has the **lowest asymptotic covariance matrix** among all possible GMM estimators

Optimal weights

If the **weighting matrix** is set to the **inverse of the long-run covariance matrix**, it puts most weight on the sample moments with lowest sampling variation, that is,

$$A_T = S^{-1},$$

where S is the long-run covariance matrix of the sample moments defined on the following slide

- Suppose moment conditions are asset pricing errors. Then this weighting matrix puts **most (least) weight** on the assets with **least (most) variance** of their pricing errors

Choice of weighting matrix

- The **long-run covariance matrix** is defined as

$$S \equiv \lim_{T\to\infty} \mathsf{Var}[\sqrt{T} g_T(\theta_0)]$$
$$= \sum_{s=-\infty}^{\infty} \mathbb{E}[G(\mathsf{data}_t, \theta_0) G(\mathsf{data}_{t-s}, \theta_0)'] \tag{38}$$

  where θ_0 is the population (true) parameters.
- If observations are independent, this reduces to

$$S = \mathbb{E}[G(\mathsf{data}_t, \theta_0) G(\mathsf{data}_t, \theta_0)']. \tag{39}$$

- The estimator of S, $\hat{S}$, **requires estimated parameters**, $\hat{\theta}$, and is, as such, infeasible at first (put “hats" on everything unknown in the equations) ...for that reason, **we need an additional step**

Two-stage and iterated GMM

Two-stage GMM

1. **Estimate GMM parameters**, using (37), with A_T equal to an arbitrary, but fixed, choice of matrix. Often, this is $A_T = I_N$. This generates $\hat{\theta}^{(1)}$, which is consistent and asymptotically normal. Use $\hat{\theta}^{(1)}$ to estimate $\hat{S}$
2. **Estimate second-stage GMM parameters**, using (37), with $A_T = \hat{S}^{-1}$. This generates $\hat{\theta}^{(2)}$, which is consistent and asymptotically normal. (Note that there is an error in Campbell (2017), as he forgets to invert $\hat{S}$ in his equation (4.88).)

- Note that the asymptotic properties are similar for each stage, yet in finite samples it is sometimes **beneficial to continue the procedure** by using $\hat{\theta}^{(2)}$ to update the estimate of $\hat{S}$ and then re-estimate parameters to get $\hat{\theta}^{(3)}, \ldots,$ until one stops when the errors, $Q(\hat{\theta}) = g_T(\hat{\theta})' A_T g_T(\hat{\theta})$, are sufficiently small

Asymptotic distribution and hypothesis testing

Asymptotic distributions for arbitrary weighting matrix

As $T \to \infty$ and any fixed A_T, it holds that

$$\hat{\theta} \xrightarrow{d} N(\theta_0, T^{-1}V),$$
$$g_T(\hat{\theta}) \xrightarrow{d} N(0, T^{-1}\Omega), \tag{40}$$

where "$\xrightarrow{d}$" means **convergence in distribution** or, loosely speaking, **"is distributed as"**, and

$$V = (D'A_TD)^{-1}D'A_TSA_T(D'A_TD)^{-1}, \tag{41}$$
$$\Omega = \left(I_N - D(D'A_TD)^{-1}D'A_T\right) S \left(I_N - A_TD(D'A_TD)^{-1}D'\right)', \tag{42}$$

and D is $\mathbb{E}[\partial g_T(\theta_0)/\partial\theta_0]$ the population first derivatives (or gradient, depending on the number of parameters)

Asymptotic distribution and hypothesis testing

Asymptotic distributions for optimal weighting matrix

As $T \to \infty$ and $A_T = S^{-1}$, it holds that

$$\begin{aligned} \hat{\theta} &\xrightarrow{d} N(\theta_0, T^{-1}V), \\ g_T(\hat{\theta}) &\xrightarrow{d} N(0, T^{-1}\Omega), \end{aligned} \tag{43}$$

and

$$V = \left(D'S^{-1}D\right)^{-1}, \tag{44}$$

$$\Omega = \left(S - D(D'S^{-1}D)^{-1}D'\right). \tag{45}$$

- Error in Campbell (2017) eq. (4.85) as he forgets to invert (44)

Hansen's J-test for overall fit

- As a test of the **overall fit** of the model, one may apply **Hansen's J-test** (also known as a test for overidentifiying restions)
- This test examines whether $g_T(\hat{\theta})$ is sufficiently close to zero

Hansen's J-test for overall fit

For the **arbitrary weighting matrix**, Hansen's J-test is defined as

$$J_T \equiv g_T(\hat{\theta})'\hat{\Omega}^+ g_T(\hat{\theta}) \xrightarrow{d} \chi^2_{N-K}, \tag{46}$$

where $\hat{\Omega}^+$ denotes the (Moore-Penrose) pseudoinverse of the (estimated) sample moment covariance matrix. (Error in Campbell (2017) eq. (4.82), missing a transpose in the first term.)

- Note that Campbell (2017) applies a quite different procedure in estimation and testing the overidentifiying restrictions

Hansen's J-test for overall fit

Hansen's J-test for overall fit

For the **optimal weighting matrix**, Hansen's J-test is defined as

$$J_T \equiv T g_T(\hat{\theta})' \hat{S}^{-1} g_T(\hat{\theta}) \xrightarrow{d} \chi^2_{N-K}. \tag{47}$$

- That is, Hansens J-test is simply T times the minimand, that is, $TQ(\hat{\theta})$ using the optimal weighting matrix
- The reduction in degrees of freedom is not of our concern here, but is the reason that we need the pseudoinverse in (46)

Inference on parameter(s)

- We can also **make hypothesis tests** on whether a **parameter** (or a group of parameters) is **equal to zero** (or something else for that matter)
- For a single, the i'th, parameter, we form a conventional t-statistic as per

$$\frac{\hat{\theta}_i}{\sqrt{\text{Var}[\hat{\theta}_{ii}]}} \xrightarrow{d} N(0,1), \tag{48}$$

  where $\text{Var}[\hat{\theta}_{ii}]$ is the i'th diagonal element of the estimate of the covariance matrix of parameters, $\hat{V}$.
- **For a group** of p many parameters, we form a **conventional Wald-type statistic** as per

$$\hat{\theta}_j' \text{Var}[\hat{\theta}_{jj}]^{-1} \hat{\theta}_j \xrightarrow{d} \chi^2_p, \tag{49}$$

  where $\hat{\theta}_j$ is a subvector of parameters and $\text{Var}[\hat{\theta}_{jj}]$ a submatrix of $\hat{V}$

Estimation of covariance matrix and D

- Regardless of the choice of weighting matrix in the estimation, we, thus, need an **estimator of the long-run covariance** to make inference (hypothesis testing)
- We also need to **estimate D**
- If the derivative is not easily obtainable in analytical form (which it is in many cases later in our lecture), numerical differentiation is easier

Estimation of covariance matrix and D

- To see the intuition, suppose θ is one-dimensional. The one-sided or forward numerical derivative is then

$$\hat{D} = \frac{g_T(\hat{\theta} + h) - g_T(\hat{\theta})}{h},$$

  where h is a very small number, e.g. $h = 1\text{e-}6$
- This is motivated from the definition of a derivative by

$$\lim_{\varepsilon \to 0} \frac{g_T(\hat{\theta} + \varepsilon) - g_T(\hat{\theta})}{\varepsilon}.$$

- We use this forward version for computational reasons
- If θ is multi-dimensional, one needs the **gradient**, and the numerical differentiation is conducted with respect to each element in θ.

Estimation of covariance matrix and D

- When estimating the **long-run covariance matrix**, S, we will distinguish between cases with or without **serial correlation**
- The first case without serial correlation can actually be motivated from the asset pricing context (see below)

Long-run covariance matrix estimation

Under **no serial correlation**, the long-run covariance matrix, S, is estimated by

$$\hat{S}(\hat{\theta}) = T^{-1} \sum_{t=1}^{T} \left(G(\text{data}_t,\hat{\theta}) - \bar{G}(\text{data}_t,\hat{\theta})\right) \times \left(G(\text{data}_t,\hat{\theta}) - \bar{G}(\text{data}_t,\hat{\theta})\right)', \tag{50}$$

where $\bar{G}(\text{data}_t,\hat{\theta}) = T^{-1} \sum_{t=1}^{T} G(\text{data}_t,\hat{\theta}) = g_T(\hat{\theta})$.

Estimation of covariance matrix and D

- If theory does not imply no serial correlation (see below), or if we want to construct tests that are robust to the presence of serial correlation, we have a **parametric** or **nonparametric** approach
- The **parametric** approach estimates a **VARMA model** for $G(\text{data}_t, \theta)$
- Alternatively, we can estimate S **nonparametrically** by a **heteroskedasticity- and autocorrelation-consistent** (HAC) covariance matrix estimator
- This is essentially a **weighted average** of all sample autocovariances of $G(\text{data}_t, \hat{\theta})$

Estimation of covariance matrix and D

HAC estimator of long-run covariance matrix

HAC estimators of the long-run covariance matrix take the form

$$\hat{S}_{HAC}(\hat{\theta}) = \hat{\Gamma}_0 + \sum_{i=1}^{T-1} \omega_i \left(\hat{\Gamma}_i + \hat{\Gamma}_i'\right), \tag{51}$$

where ω_i is a kernel (or weight), and

$$\begin{aligned}\hat{\Gamma}_i = T^{-1} \sum_{t=i+1}^{T} & \left(G(\text{data}_t,\hat{\theta}) - \bar{G}(\text{data}_t,\hat{\theta})\right) \\ & \times \left(G(\text{data}_{t-i},\hat{\theta}) - \bar{G}(\text{data}_{t-i},\hat{\theta})\right)' \end{aligned} \tag{52}$$

is the i'th sample autocovariance matrix

Estimation of covariance matrix and D

- Higher-order autocovariances need to be down-weighted to ensure consistency and positive semi-definiteness in all (finite) samples
- A common kernel choice is the Bartlett kernel by Newey and West (1987), given by

$$\omega_i = \begin{cases} 1 - \frac{i}{m+1}, & \text{for } i \leq m+1, \\ 0, & \text{for } i \geq m+1, \end{cases}$$

  where $m \geq 0, m \in \mathbb{Z}$, is the bandwidth that controls the number of autocovariances included in the estimator.
- In practice, one needs to make sure that the choice of m does not leave out important autocovariances
- …by, e.g., trying different candidate values and ensuring that adding additional autocovariances will not affect the HAC estimate significantly (or picking it optimally and data-driven (Andrews, 1991))

Asset pricing meets GMM

Asset pricing meeting GMM

- While the **moment conditions** in GMM are all **unconditional** in the presentation so far, most **asset pricing models** imply results for **conditional moments** (e.g. the CCAPM), as per

$$\mathbb{E}[G(\text{data}_t,\theta)|\mathcal{F}_t] = 0. \tag{53}$$

- This essentially requires explicit modelling of the conditional distributions, which is often complicated
- Rather, we can focus on the **implications** for **unconditional models** derived from **conditional models** and test those

Asset pricing meeting GMM

- An **asset pricing model** expressed in **conditional moments** implies **two sets of unconditional moment constraints**:
  - A conditioning down principle
  - Instruments that stand in for conditioning information in $\mathcal{F}_t$

Asset pricing meeting GMM

Implication 1: Conditioning down

Taking **unconditional expectations** of (21) and using the **law of iterated expectations** yields

$$\begin{aligned} \mathbb{E}[P_{it}] &= \mathbb{E}\left[\mathbb{E}_t[M_{t+1}X_{it+1}]\right] \\ &= \mathbb{E}[M_{t+1}X_{it+1}]. \end{aligned} \tag{54}$$

- This has a similar structure as the conditional expression, yet the implied moment condition is

$$\mathbb{E}[M_{t+1}X_{it+1} - P_{it}] = 0, \tag{55}$$

with $G(\cdot) = M_{t+1}X_{it+1} - P_{it}$.

Asset pricing meeting GMM

- Let z_t be a so-called **instrument** observed at time t. For any random variable y_{t+1} it can be shown that, if

$$\mathbb{E}[y_{t+1}z_t] = 0, \quad \forall z_t \in \mathcal{F}_t, \tag{56}$$

  then it implies

$$\mathbb{E}[y_{t+1}|\mathcal{F}_t] = 0. \tag{57}$$

- Setting $y_{t+1} = M_{t+1}X_{it+1} - P_{it}$ reveals that

$$\mathbb{E}[(M_{t+1}X_{it+1} - P_{it})z_t] = 0, \quad \forall z_t \in \mathcal{F}_t, \tag{58}$$

  is **sufficient** for **estimating/testing the conditional model** of $P_{it} = \mathbb{E}_t[M_{t+1}X_{it+1}]$

Asset pricing meeting GMM

- Start with the **fundamental pricing equation** in (21) and multiply an instrument to get

$$P_{it} z_t = \mathbb{E}_t[M_{t+1} X_{it+1} z_t], \tag{59}$$

where z_t can “move freely" in and out of expectations as it is adapted to $\mathcal{F}_t$ (known at time t).

Implication 2: Scaled payoffs

Unconditional expectations and the law of iterated expectations yield

$$\mathbb{E}[P_{it} z_t] = \mathbb{E}[M_{t+1} X_{it+1} z_t]. \tag{60}$$

- Doing this for all z_t **generates a set of implications** **not captured by Implication 1**

Asset pricing meeting GMM

- The moment conditions implied are thus

$$\mathbb{E}[(M_{t+1}X_{it+1} - P_{it})z_t] = 0, \tag{61}$$

  where $G(\cdot) = (M_{t+1}X_{it+1} - P_{it})z_t$.
- In practice, we of course have to choose a limited set of instruments ➨ natural source of critique.
- It can be understood in the context of scaled payoffs and managed portfolios
- $\tilde{X}_{it+1} = X_{it+1}z_t$ is an alternative asset with a scaled payoff and it has price $\tilde{P}_{it} = P_{it}z_t$. Here, z_t is a weighting variable, that informs the manager/investor on how much to buy or sell of a given asset
- For instance, high z_t can be informative/forecast high returns and he/she should buy more and vice versa. As such, z_t scales the investment, naturally scaling the payoff and the price

Asset pricing meeting GMM

- Using Implication 1 and 2 is, in principle, sufficient for capturing all unconditional implications of the conditional model

Implications 1 and 2 in return form

We will almost always work with returns (to ensure stationary data) and the resulting moment conditions are for each return

$$\text{Implication} \quad 1: \mathbb{E}[M_{t+1}R_{it+1} - 1] = 0, \tag{62}$$

$$\text{Implication} \quad 2: \mathbb{E}[(M_{t+1}R_{it+1} - 1)z_t] = 0, \quad \forall z_t \in \mathcal{F}_t. \tag{63}$$

- Suppose we have several, e.g. n many, tests asset ➨ we then have a vector returns
- Denote this by $R_{t+1} = (R_{1t+1}, \ldots, R_{nt+1})'$.

Asset pricing meeting GMM

- Moreover, suppose we have q many instruments (excluding the constant), which we gather in a vector $Z_t = (1, z_{1t}, \dots, z_{q,t})'$ (that includes the constant)
- We can then **express both implications** compactly in a **single equation**

Kronecker formulation of Implications 1 and 2

Implications 1 and 2, using R_{t+1} and Z_t, reads

$$\mathbb{E}[(M_{t+1}R_{it+1} - 1) \otimes Z_t] = 0, \tag{64}$$

where "$\otimes$" is the **Kronecker/tensor product** and means "multiply every element by every other element". This leads to $n(q+1)$ moment conditions

Asset pricing meeting GMM

Example: Kronecker formulation

Suppose $Z_t = (1,z_{1t})'$ and $R_{t+1} = (R_{1t+1},R_{2t+1})'$. Then (64) is

$$\mathbb{E}\left[\begin{pmatrix} M_{t+1}R_{1t+1} \\ M_{t+1}R_{2t+1} \\ M_{t+1}R_{1t+1}z_{1t} \\ M_{t+1}R_{2t+1}z_{1t} \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \\ z_{1t} \\ z_{1t} \end{pmatrix}\right] = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \tag{65}$$

yielding $n(q+1) = 2(1+1) = 4$ moment conditions

Asset pricing meeting GMM: Summary

- The **asset pricing model** says that although conditional expected returns can vary over time, **discounted returns** should **always be the same, 1**
- The model prediction error is $U_{it+1} \equiv M_{t+1}R_{it+1} - 1$. The asset pricing model says it should be both conditionally and unconditionally zero (Implication 1)
- If the asset pricing model is supposed to be **true**, we should not be able to use **any information today**, i.e., z_t, to **forecast any of the errors** (Implication 2). This is exactly what testing $\mathbb{E}[U_{it+1}z_t] = 0$ means

Asset pricing meeting GMM

No serial correlation as per asset pricing models

Recall that the asset pricing models delineate that $\mathbb{E}[U_{t+1}z_t] = 0$ and $\mathbb{E}[U_{t+1}] = 0$. This also means that these conditions imply that if $z_t = U_t$, it has to satisfy $\mathbb{E}[U_{t+1}U_t] = 0$. That is, **they imply no serial correlation** since

$$\text{Cov}[U_{t+1}, U_t] = \mathbb{E}[U_{t+1}U_t] - \mathbb{E}[U_{t+1}]\mathbb{E}[U_t] = \mathbb{E}[U_{t+1}U_t] = 0.$$

- …but **use the HAC covariance matrix anyway**, since no asset pricing model is really the true one!

Choosing weighting matrix A_T

- Recall that the **choice of weighting matrix** is essentially a choice on **how to weight the sample moments** in the GMM estimation
- This mostly comes down to choosing between setting $A_T = I_N$ or $A_T = S^{-1}$
- In the **context of asset pricing**, the **former weights all pricing errors equally** among assets, whereas the latter puts more emphasis on those **assets that are most precisely predicted**

Choosing weighting matrix A_T

1. If a **single model** is estimated and inference on its asset pricing ability is made only on this model, it is recommended to use the optimal weighting matrix $A_T = S^{-1}$
2. If a **pair or several models** are estimated and their asset pricing abilities compared, it is recommend to use the identity weighting matrix $A_T = I_N$

Choosing weighting matrix A_T

- Since S^{-1} (most likely) changes according to the model, one model may "improve" $J_T \equiv g_T(\hat{\theta})'\hat{\Omega}^+ g_T(\hat{\theta})$ simply because it blows up S rather than making the pricing errors smaller
- Moreover, **if the risk-free rate** is included as a test asset (which it typically is), then J_T essentially evaluates how well each model prices the Risk-free bond if S^{-1} is used, **ignoring all the other assets**
- As such, **one has to use a common weighting matrix across all models** to answer whether one model leads to s**maller pricing errors** (describes data better) than others

Example

Example: Power utility CCAPM

- Let us consider an example in the Matlab live script *GMM_SDF.mlx*.

References

ANDREWS, D. W. (1991): "Heteroskedasticity and autocorrelation consistent covariance matrix estimation," *Econometrica*, 817–858.

CAMPBELL, J. Y. (2017): *Financial decisions and markets: a course in asset pricing*, Princeton University Press.

COCHRANE, J. H. (2009): *Asset pricing: Revised edition*, Princeton university press.

NEWEY, W. K. AND K. D. WEST (1987): "A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix," *Econometrica*, 55, 703–708.