Data Analysis: Estimation

Mark Hendricks

August Review

UChicago Financial Mathematics

Outline

The Projection Estimate

Standard errors

Heteroskedasticity

Estimators

- Estimators are functions of random data.
- Thus, the output of this function, the estimate, is itself a random variable.
- Its randomness comes from the randomness of the function's input.

For example,

- $\overline{x}$ is a function of sample data $\{x_i\}$.
- Random variation in the different samples will cause random variation in the sample average, $\overline{x}$.
- Note that in some samples, the resulting estimate is far from the true parameter.

Questions for an estimator

- What estimator, (function,) is being used?
- Do the random estimates center around the true parameter?
- Do the random estimates tightly cluster around the true parameter?
- What is the distribution of the estimates? i.e. What is the probability that the estimate is a certain distance from the true parameter value?

Poor estimator

Example (Lazy estimator)

Consider again the lazy estimator of the mean of x_i:

$$\hat{x} = x_1$$

Figure 1 shows that this estimator does not center nor bunch tightly around μ.

Illustration of randomness

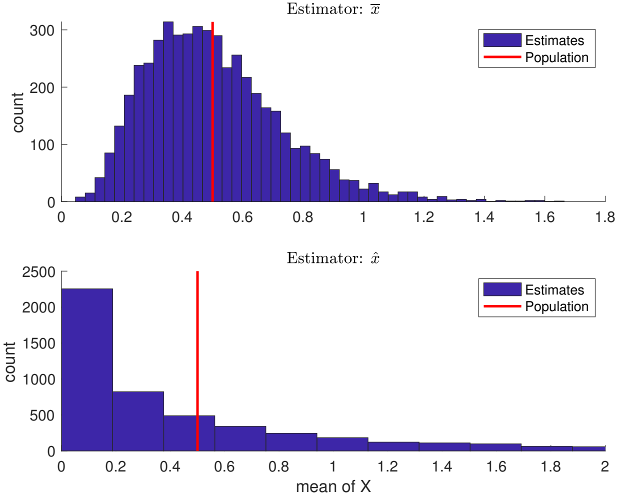

Figure: Data is simulated from a gamma distribution, with shape 0.5 and scale 1.0. The histogram is based on 5000 estimates of $\overline{x}$, which each are estimated from a sample size of 10.

Sample notation

Suppose we observe n realizations of the stochastic process, $\{y_i, \mathbf{x}_i\}$.

- The observations are denoted $(y_i, \mathbf{x}_i)$ for $i = 1 \ldots n$.
- The total sample is denoted with the $n \times 1$ vector $\mathbf{y}$ and the $n \times k$ vector $\mathbf{X}$, where each row of $\mathbf{X}$ is an observation, $\mathbf{x}_i'$.

$$\underbrace{\mathbf{Y}}_{n \times 1} = \underbrace{\mathbf{X}}_{n \times k} \underbrace{\vec{b}}_{k \times 1} + \underbrace{\vec{e}}_{n \times 1}$$

Sample vs population errors

Consider estimating the projection from this sample:

- An estimate of β with some coefficient vector, $\vec{b}$, will decompose the sample, $\{y, \mathbf{x}\}_n$ into a sample of $\mathbf{x}'\vec{b}$ and $\vec{e}$.
- The projection errors, ϵ_i, associated with the sample values will be unobserved and unknown.

Notation example

Consider the sample notation applied to the example from prior,

$$\underbrace{\begin{bmatrix} \text{mean return}_1 \\ \text{mean return}_2 \\ \vdots \\ \text{mean return}_n \end{bmatrix}}_{\mathbf{Y}} = \underbrace{\begin{bmatrix} \text{div}_1 & \text{vol}_1 \\ \text{div}_2 & \text{vol}_2 \\ \vdots & \vdots \\ \text{div}_n & \text{vol}_n \end{bmatrix}}_{\mathbf{X}} \underbrace{\begin{bmatrix} -.0009 \\ .0796 \end{bmatrix}}_{\boldsymbol{\beta}} + \underbrace{\begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}}_{\text{unobserved error}}$$

Contrast this with the notation for a single observation:

$$\underbrace{\text{mean return}_i}_{y_i} = \underbrace{\begin{bmatrix} \text{div}_i & \text{vol}_i \end{bmatrix}}_{(\mathbf{x}_i)'} \underbrace{\begin{bmatrix} -.0009 \\ .0796 \end{bmatrix}}_{\boldsymbol{\beta}} + \epsilon_i$$

Projection estimate

Estimate the projection coefficient vector, $\mathbf{b}$, using the same formula but simply replacing the population moments with their sample averages

$$\mathbf{b} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y}$$

- We must assume $\mathbf{X}'\mathbf{X}$ is nonsingular—that it has a non-zero determinant.
- This assumption will rarely be a problem mathematically.
- We will find that "nearly" violating it causes our biggest statistical issues.

Sample decomposition

This estimator ensures the resulting in-sample decomposition is indeed a projection,

$$\begin{aligned} \mathbf{Y} &= \mathbf{X}\mathbf{b} + \vec{e} \\ \mathbf{Y} &= \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y} + \vec{e} \\ \mathbf{Y} &= \mathbf{P}\mathbf{Y} + \vec{e} \end{aligned}$$

where the projection matrix, $\mathbf{P}$ is

$$\mathbf{P} = \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'$$

$$\mathbf{P}^2 = \mathbf{P} \qquad \mathbf{P}\vec{e} = 0 \qquad \mathbf{X}'\vec{e} = 0$$

b as a random variable

The estimated linear projection vector is itself a random variable, the sum of the population projection vector plus an error term:

$$\begin{aligned}\mathbf{b} &= \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y} \\ &= \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\left(\mathbf{X}\beta + \epsilon\right) \\ &= \underbrace{\beta}_{\text{constant}} + \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\epsilon \qquad (1)\end{aligned}$$

Thus, the estimate is a constant plus a random variable estimation error:

$$\underbrace{\mathbf{b}}_{\text{random vector}} = \underbrace{\beta}_{\text{unknown constant vector}} + \underbrace{\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\epsilon}_{\text{estimation error}}$$

Sample versus population

- $\boldsymbol{\epsilon}$ is the $n \times 1$ vector of projection errors associated with the sample data, $\{y_i, \mathbf{x}_i\}$. They are unobserved.
- $\vec{e}$ is the $n \times 1$ vector of estimated projection errors associated with the sample data and the sample estimate, **b**.
- By construction, $\boldsymbol{\epsilon}$ is orthogonal to **x** in population, while $\vec{e}$ is orthogonal to the sample **X**.
- Even though ϵ is orthogonal to **x** in population, it almost surely will not be orthogonal to the sample **X**, due to random variation.
- The estimation error, $\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\epsilon}$ is an unobserved random variable, given that it is the function of the unobserved $n \times 1$ vector, $\boldsymbol{\epsilon}$.

Consistent projection

Theorem

Assuming identification and stationary ergodicity, the estimator **b** is a consistent estimator of $\boldsymbol{\beta}$.

$$\mathbf{b} \xrightarrow{P} \boldsymbol{\beta}$$

In fact, we can view **b** as being based on consistent estimates for the two second moments in a projection:

$$\frac{1}{n}\left(\mathbf{X}'\mathbf{X}\right) \xrightarrow{P} \mathbb{E}\left[\mathbf{x}\mathbf{x}'\right]$$

$$\frac{1}{n}\left(\mathbf{X}'\mathbf{Y}\right) \xrightarrow{P} \mathbb{E}\left[\mathbf{x}y\right]$$

$$\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y} \xrightarrow{P} \left(\mathbb{E}\left[\mathbf{x}\mathbf{x}'\right]\right)^{-1}\mathbb{E}\left[\mathbf{x}y\right] = \boldsymbol{\beta}$$

Variability of the estimation error

Consider the second moment of the estimation error.

$$\boldsymbol{\nu}_i \equiv \mathbf{x}_i \epsilon_i$$

Define the sample averages estimating second moments,

$$\mathbf{S}_{\mathbf{x}} \equiv \frac{1}{n}\mathbf{X}'\mathbf{X} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i'$$

$$\bar{\boldsymbol{\nu}} \equiv \frac{1}{n}\mathbf{X}'\boldsymbol{\epsilon} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i \epsilon_i$$

We can write **b** as,

$$\mathbf{b} = \boldsymbol{\beta} + \mathbf{S}_{\mathbf{x}}^{-1}\bar{\boldsymbol{\nu}}$$

Formula for covariation in **b**

Then the variation in **b** is

$$\begin{aligned}\Sigma_{\mathbf{b}} =& \mathbb{E}\left[\left(\mathbf{b}-\beta\right)\left(\mathbf{b}-\beta\right)' \mid \mathbf{X}\right] \\ =& \mathbb{E}\left[\mathbf{S}_{\mathbf{x}}^{-1}\bar{\boldsymbol{\nu}}\bar{\boldsymbol{\nu}}'\mathbf{S}_{\mathbf{x}}^{-1} \mid \mathbf{X}\right]\end{aligned}$$

Rewrite this as

$$\begin{aligned}\Sigma_{\mathbf{b}} =& \mathbf{S}_{\mathbf{x}}^{-1}\Sigma_{\boldsymbol{\nu}}\mathbf{S}_{\mathbf{x}}^{-1} \\ \Sigma_{\boldsymbol{\nu}} \equiv& \ \mathbb{E}\left[\bar{\boldsymbol{\nu}}\bar{\boldsymbol{\nu}}' \mid \mathbf{X}\right]\end{aligned} \tag{2}$$

Asymptotic distribution of **b**

Theorem (Limiting distribution of **b**)

Assuming identification and stationary ergodicity,

$$\sqrt{n}\left(\mathbf{b}-\beta\right) \xrightarrow{D} \mathcal{N}\left(\mathbf{0}, \Sigma_{\mathbf{b}}^{\text{lim}}\right)$$

where we define notation,

$$\begin{aligned} \Sigma_{\mathbf{x}} &\equiv \mathbb{E}\left[\mathbf{x}\mathbf{x}'\right] \\ \Sigma_{\mathbf{b}}^{\text{lim}} &\equiv \Sigma_{\mathbf{x}}^{-1}\Sigma_{\nu}\Sigma_{\mathbf{x}}^{-1} \end{aligned}$$

Observations

- Note that the asymptotic covariance is the same as the finite sample covariance in (2), but replacing the sample estimates $\mathbf{S_x}$ with the population moment $\mathbb{E}[\mathbf{xx}']$.
- We refer to $\Sigma_{\mathbf{b}}^{\lim}$ as the asymptotic covariance matrix of $\mathbf{b}$.
- The theorem is important as it tells us how the random estimate, $\mathbf{b}$ is distributed around the value we are estimating, $\boldsymbol{\beta}$. This is necessary in order to test hypotheses on $\boldsymbol{\beta}$

Outline

The Projection Estimate

Standard errors

Heteroskedasticity

Errors that are i.i.d.

Let $\mathcal{I}$ denote the identity matrix. Assuming i.i.d.data,

$$\mathbb{E}\left[\epsilon_i \epsilon_j \,|\, \mathbf{x}\right] =0,\ i \neq j$$
$$\mathbb{E}\left[\epsilon_i \epsilon_i \,|\, \mathbf{x}\right] =\gamma_0^2,\ i = 1, \ldots, n$$

This simplifies Equation (2),

$$\Sigma_{\boldsymbol{\nu}} = \mathbb{E}\left[\mathbf{x}_i \epsilon_i \epsilon_i \mathbf{x}_i'\right]$$
$$=\Sigma_{\mathbf{x}} \Sigma_{\epsilon}$$
$$\Sigma_{\epsilon} \equiv \mathbb{E}\left[\boldsymbol{\epsilon}\boldsymbol{\epsilon}' \,|\, \mathbf{x}\right] = \gamma_0^2 \mathcal{I}$$

Standard errors under i.i.d.

Thus,

$$\Sigma_{\mathbf{b}} = \gamma_0^2 \; \Sigma_{\mathbf{x}}^{-1}$$

Estimate these moments with the sample averages,

$$\mathbf{s}_0^2 \equiv \frac{1}{n} \vec{e}' \vec{e}$$

$$\mathbf{S}_{\mathbf{b}} = \mathbf{s}_0^2 \; \left(\mathbf{X}'\mathbf{X}\right)^{-1} \tag{3}$$

Homoskedasticity

Homoskedasticity refers to having constant second moments.

$$\mathbb{E}\left[\epsilon_i \epsilon_j\right] = \sigma_{i,j}$$

Assuming stationarity, $\{\epsilon_i\}$ is unconditionally homoskedastic:

$$\mathbb{E}\left[\epsilon_i \epsilon_j\right] = \gamma_{i-j}$$

Conditional homoskedasticity

For estimation of $\Sigma_{\boldsymbol{\nu}}$, we are interested in whether $\{\epsilon_i\}$ is conditionally homoskedastic, which is not guaranteed by stationarity ergodicty.

Assumption (Conditional Homoskedasticity)

The process $\{\epsilon_i\}$ is homoskedastic conditional on $\{\mathbf{x}_i\}$.

$$\mathbb{E}\left[\epsilon_i \epsilon_j \,\middle|\, \mathbf{x}_i, \mathbf{x}_j\right] = \sigma_{i,j}$$

Standard errors

Under Conditional Homoskedasticity, $\Sigma_{\boldsymbol{\nu}}$ simplifies to

$$\Sigma_{\boldsymbol{\nu}} = \mathbf{X}'\Sigma_{\epsilon}\mathbf{X}$$

- Thus, ϵ_i and ϵ_j can covary, but this covariance cannot depend on $\mathbf{x}$.
- To directly estimate Σ_{ϵ}, we have only n realizations with which to estimate the $n(n+1)/2$ unique elements.
- We need to add more structure in order to get any power to this estimate.

Stationary covariance

Under stationarity, the joint distribution of (ϵ_i, ϵ_j) depends only on $h = i - j$, not on i, j. Then for any pair of points in the sample,

$$\mathbb{E}\left[\epsilon_i \epsilon_j\right] = \mathbb{E}\left[\epsilon_{i+h} \epsilon_{j+h}\right]$$

This restricts the covariance matrix of n realizations of $\{\epsilon_i\}$ to

$$\Sigma_\epsilon = \begin{bmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \cdots & \gamma_{n-3} & \gamma_{n-2} & \gamma_{n-1} \\ \gamma_1 & \gamma_0 & \gamma_1 & \cdots & \gamma_{n-4} & \gamma_{n-3} & \gamma_{n-2} \\ \gamma_2 & \gamma_1 & \gamma_0 & \cdots & \gamma_{n-5} & \gamma_{n-4} & \gamma_{n-3} \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ \gamma_{n-3} & \gamma_{n-4} & \gamma_{n-5} & \cdots & \gamma_0 & \gamma_1 & \gamma_2 \\ \gamma_{n-2} & \gamma_{n-3} & \gamma_{n-4} & \cdots & \gamma_1 & \gamma_0 & \gamma_1 \\ \gamma_{n-1} & \gamma_{n-2} & \gamma_{n-3} & \cdots & \gamma_2 & \gamma_1 & \gamma_0 \end{bmatrix}$$

Estimating Σ_ϵ

Given a sample size of n, we can then estimate each γh with $\mathbf{s}h$:

$$\mathbf{s}h = \frac{1}{n-h} \sum_{i=h+1}^{n} e_i e_{i-h} \tag{4}$$

Let $\mathbf{S}_e$ denote the $n \times n$ matrix of these estimates, $\mathbf{s}h$, for $0 \leq h \leq n-1$. Then use the estimators,

$$\begin{aligned} \mathbf{S}_{\boldsymbol{u}} &= \mathbf{X}'\mathbf{S}_e\mathbf{X} \\ \mathbf{S}_{\mathbf{b}} &= \mathbf{S}_{\mathbf{x}}^{-1}\mathbf{S}_{\boldsymbol{u}}\mathbf{S}_{\mathbf{x}}^{-1} \end{aligned} \tag{5}$$

Restricting high-order correlation

Assumption (Orthogonality)

Error terms, ϵ_i and ϵ_{i-h} are conditionally orthogonal beyond for large h.

$$\mathbb{E}\left[\epsilon_i \epsilon_j\right] = 0, \quad \forall \ |i-j| \ > H$$

Suppose this assumption holds for $H = 2$,

$$\Sigma_\epsilon = \begin{bmatrix} \gamma 0 & \gamma 1 & \gamma 2 & \dots & & & \\ \gamma 1 & \gamma 0 & \gamma 1 & \dots & & \mathbf{0} & \\ \gamma 2 & \gamma 1 & \gamma 0 & \dots & & & \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ & & & \dots & \gamma 0 & \gamma 1 & \gamma 2 \\ & \mathbf{0} & & \dots & \gamma 1 & \gamma 0 & \gamma 1 \\ & & & \dots & \gamma 2 & \gamma 1 & \gamma 0 \end{bmatrix}$$

Outline

The Projection Estimate

Standard errors

Heteroskedasticity

The general case

Recall

$$\boldsymbol{\nu}_i \equiv \mathbf{x}_i \epsilon_i$$

$$\begin{aligned}\Sigma_{\mathbf{b}} =& \mathbf{S}_{\mathbf{x}}^{-1} \Sigma_{\boldsymbol{\nu}} \mathbf{S}_{\mathbf{x}}^{-1} \\ \Sigma_{\boldsymbol{\nu}} \equiv& \mathbb{E}\left[\bar{\boldsymbol{\nu}}\bar{\boldsymbol{\nu}}' \,\middle|\, \mathbf{X}\right]\end{aligned}$$

Note that $\Sigma_{\boldsymbol{\nu}}$ is the second moment matrix of a sample average, $\bar{\boldsymbol{\nu}}$. Thus, apply the general CLT.

The general equation

Equation (6) indicates that in general,

$$
\begin{aligned}
\Sigma_{\nu} &= \frac{1}{n^2} \sum_{h=-(n-1)}^{n-1} (n - |h|)\, \Gamma_h^{\nu} \\
&= \frac{1}{n^2} \left[n\, \Gamma_0^{\nu} + 2 \sum_{h=1}^{n-1} (n - h) \left(\Gamma_h^{\nu} + (\Gamma_h^{\nu})' \right) \right] \\
\Gamma_h^{\nu} &\equiv \mathbb{E}\left[\boldsymbol{\nu}_i \boldsymbol{\nu}_{i-h}' \right]
\end{aligned}
$$

Estimating all orders

This suggests an estimator for $\mathbf{S}_{\boldsymbol{u}}$, with elements that use sample averages to replace the covariance moments above.

$$\mathbf{S}_{\boldsymbol{u}} = \frac{1}{n^2} \sum_{h=-(n-1)}^{n-1} (n - |h|)\, \mathcal{S}_h$$
$$= \frac{1}{n^2} \left[n\,\mathcal{S}_0 + \sum_{h=1}^{n-1} (n-h) \left(\mathcal{S}_h + \mathcal{S}_h' \right) \right]$$

Define the estimator of the mixed second-moments,

$$\mathcal{S}_h \equiv \frac{1}{n-h} \sum_{i=h+1}^{n} \boldsymbol{u}_i \boldsymbol{u}_{i-h}'$$

The sample estimator

Substituting, we have the well-known spectral density estimate or $\Sigma_{\boldsymbol{\nu}}$,

$$\mathbf{S}_{\boldsymbol{u}} = \frac{1}{n^2}\left[\sum_{i=1}^{n} \boldsymbol{u}_i \boldsymbol{u}_i' + \sum_{h=1}^{n-1}\sum_{i=h+1}^{n}\left(\boldsymbol{u}_i \boldsymbol{u}_{i-h}' + \boldsymbol{u}_{i-h}\boldsymbol{u}_i'\right)\right] \tag{6}$$

Restricting for large h

The Hansen-Hodrick correction implements Equation (6) with additionally making Assumption 2.

$$\mathcal{S}_h = \begin{cases} \dfrac{1}{n-h} \displaystyle\sum_{i=h+1}^{n} \boldsymbol{u}_i \boldsymbol{u}_{i-h}' & \text{for } h < H \\ \mathbf{0} & \text{for } h \geq H \end{cases}$$

Thus the restriction limits the estimate to considering H total sample second-moments,

$$\mathbf{S}_{\boldsymbol{u}} = \frac{1}{n^2} \left[\sum_{i=1}^{n} \boldsymbol{u}_i \boldsymbol{u}_i' + \sum_{h=1}^{H-1} \sum_{i=h+1}^{n} \left(\boldsymbol{u}_i \boldsymbol{u}_{i-h}' + \boldsymbol{u}_{i-h} \boldsymbol{u}_i' \right) \right] \tag{7}$$

Positive semi-definiteness

Unfortunately, Equation (7) is not guaranteed to be positive definite, and often is not.

The Newey-West estimator remedies this by putting less weight on $\mathcal{S}$ for covariances with larger lags, h:

$$\mathbf{S}_{\boldsymbol{u}} = \frac{1}{n^2}\left[n\,\mathcal{S}_0 + \sum_{h=1}^{H-1}(n-h)w_h\left(\mathcal{S}_h + \mathcal{S}_h'\right)\right]$$

$$w_h \equiv 1 - \frac{h}{H}$$

Newey-West

$$\mathbf{S}_{\boldsymbol{u}} = \frac{1}{n^2}\left[\sum_{i=1}^{n} \boldsymbol{u}_i \boldsymbol{u}_i' + \sum_{h=1}^{H-1}\left(1-\frac{h}{H}\right)\sum_{i=h+1}^{n}\left(\boldsymbol{u}_i \boldsymbol{u}_{i-h}' + \boldsymbol{u}_{i-h}\boldsymbol{u}_i'\right)\right] \tag{8}$$

The Newey West Estimator in (8) is widely implemented in computational statistics libraries.

Putting it together to get $\Sigma_{\mathbf{b}}$

Using the estimators in Equations (6), (7), or (8), we finally have a consistent estimator for $\Sigma_{\mathbf{b}}$:

$$\begin{aligned}\Sigma_{\mathbf{b}} =& \mathbf{S}_{\mathbf{x}}^{-1}\mathbf{S}_{\boldsymbol{u}}\mathbf{S}_{\mathbf{x}}^{-1} \\ =& (\mathbf{XX})^{-1}\left[\sum_{i=1}^{n}\boldsymbol{u}_i\boldsymbol{u}_i' + \sum_{h=1}^{n-1}\sum_{i=h+1}^{n}\left(\boldsymbol{u}_i\boldsymbol{u}_{i-h}' + \boldsymbol{u}_{i-h}\boldsymbol{u}_i'\right)\right](\mathbf{XX})^{-1}\end{aligned}$$

where we note that the $\frac{1}{n^2}$ in $\mathbf{S}_{\boldsymbol{u}}$ cancels with the $\frac{1}{n}$ inside each $\mathbf{S}_{\mathbf{x}}$.

Estimating the asymptotic variance

- Theorem 2 contains an expression for the asymptotic covariance of $\mathbf{b}$, denoted $\Sigma_{\mathbf{b}}^{\text{lim}}$.
- Estimating this asymptotic covariance given a sample size of n is done by estimating $\Sigma_{\mathbf{b}}$ using Newey-West, conditional homoskedasticity, or i.i.d.
- As a corollary, the consistency of $\mathbf{b}$ ensures the consistency of using sample averages to estimate $\Sigma_{\mathbf{x}}, \Sigma_{\boldsymbol{\nu}}$, and $\Sigma_{\mathbf{b}}$.

$$\mathbf{s}_0^2 \xrightarrow{P} \gamma_0^2 \qquad \mathbf{S}_e \xrightarrow{P} \Sigma_\epsilon \qquad \mathbf{S}_{\boldsymbol{u}} \xrightarrow{P} \Sigma_{\boldsymbol{\nu}} \qquad \mathbf{S}_{\mathbf{b}} \xrightarrow{P} \Sigma_{\mathbf{b}}$$