7

Estimating the distribution of the market invariants

The classical approach to estimation discussed in the second part of the book does not account for estimation risk. To tackle this issue, in this chapter we introduce the Bayesian approach to parameter estimation.

The outcome of the Bayesian estimation process is the posterior distribution of the market parameters. This distribution explicitly acknowledges that an estimate cannot be a single number. Furthermore the posterior distribution includes within a sound statistical framework both the investor's experience, or prior knowledge, and the information from the market.

In Section 7.1 we introduce the Bayesian approach in general, showing how to blend the investor's prior and the market information to obtain the posterior distribution. Furthermore, we show how the Bayesian framework includes the classical approach to estimation in the form of "classical-equivalent" estimators. Finally, we discuss how to summarize the main features of a generic posterior distribution by means of its location-dispersion ellipsoid.

As in the classical approach in Chapter 4, we then proceed to discuss the estimation of the parameters of the market invariants that are most relevant to allocation problems, namely location, dispersion, and factor loadings.

In Section 7.2 we compute the posterior distribution of expected value and covariance matrix of the market invariants under the conjugate normal-inverse-Wishart hypothesis. Then we compute the classical-equivalent estimators of the above parameters, exploring their self-adjusting behavior. Finally we compute the joint and the marginal location-dispersion ellipsoids of expected values and covariance matrix provided by their posterior distribution.

In Section 7.3 we consider multivariate factor models. First we compute the posterior distribution of the factor loadings and of the perturbation covariance under the conjugate normal-inverse-Wishart hypothesis. Then we compute the classical-equivalent estimators of the respective parameters, exploring their self-adjusting behavior. Finally we compute the joint and the marginal location-dispersion ellipsoids of the posterior distribution of factor loadings and perturbation covariance.

In Section 7.4 we discuss how to quantify the investor's prior knowledge of the market in practice. Indeed, in typical situations the investor does not input directly the prior market parameters: instead, he computes them under suitable assumptions from what he considers an ideal allocation.

7.1 Bayesian estimation

We recall from Chapter 4 that a classical estimator is a function that processes current information i_T and outputs an S-dimensional vector $\widehat{\boldsymbol{\theta}}$, see (4.9). Information consists as in (4.8) of a time series of T past observations of the market invariants:

$$i_T \equiv \{\mathbf{x}_1, \ldots, \mathbf{x}_T\} . \tag{7.1}$$

The output $\widehat{\boldsymbol{\theta}}$ is a *number* which is supposed to be close to the true, unknown parameter $\boldsymbol{\theta}^{\mathrm{t}}$. We can summarize the classical approach as follows:

$$\boxed{\textbf{classical estimation: } i_T \mapsto \widehat{\boldsymbol{\theta}}} \tag{7.2}$$

The Bayesian estimation process differs from the classical one in terms of both "input" and "output".

7.1.1 Bayesian posterior distribution

In the first place, in a Bayesian context an estimator does not yield a number $\widehat{\boldsymbol{\theta}}$. Instead, it yields a random variable $\boldsymbol{\theta}$, which takes values in a given range $\boldsymbol{\Theta}$. The distribution of $\boldsymbol{\theta}$ is called the *posterior distribution*, which can be represented for instance in terms of its probability density function $f_{\mathrm{po}}(\boldsymbol{\theta})$. The true, unknown parameter $\boldsymbol{\theta}^{\mathrm{t}}$ is assumed to be hidden most likely in the neighborhood of those values where the posterior distribution is more peaked, but the possibility that $\boldsymbol{\theta}^{\mathrm{t}}$ might lie in some other region of the range $\boldsymbol{\Theta}$ is also acknowledged, see Figure 7.2.

Secondly, in a Bayesian context an estimator does not depend only on backward-looking historical information i_T. Indeed, the investor/statistician typically has some prior knowledge of the unknown value $\boldsymbol{\theta}^{\mathrm{t}}$ based on his experience e_C, where C denotes the level of confidence in his experience. This experience is explicitly taken into account in the Bayesian estimation process.

Therefore we can summarize the Bayesian approach as follows:

$$\boxed{\textbf{Bayesian estimation: } i_T, e_C \mapsto f_{\mathrm{po}}(\boldsymbol{\theta})} \tag{7.3}$$

The Bayesian approach to estimation can be interpreted intuitively as follows, see Figure 7.1.

On the one hand, the purely classical estimator based on historical information i_T gives rise to a distribution of the market parameters $\boldsymbol{\theta}$ that is

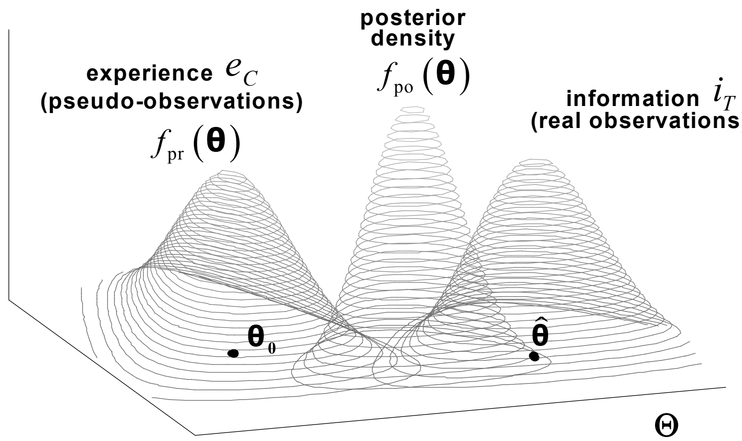

Fig. 7.1. Bayesian approach to parameter estimation

peaked around the classical estimate $\widehat{\boldsymbol{\theta}}$: the larger the number of observations T in the time series, the higher the concentration of the historical distribution around the classical estimate.

On the other hand, the investor equates his experience e_C to a number C of pseudo-observations, that only he sees, located in a "prior" value $\boldsymbol{\theta}_0$. These observations give rise to a distribution of the market parameters $\boldsymbol{\theta}$ which is called the *prior distribution*, whose probability density function we denote as $f_{\text{pr}}(\boldsymbol{\theta})$. The larger the number of these pseudo-observations, the higher the investor's confidence in his own experience and thus the more concentrated the prior distribution around $\boldsymbol{\theta}_0$.

The Bayesian posterior provides a theoretically sound way to blend the above two distributions into a third distribution, i.e. a spectrum of values and respective probabilities for the parameters $\boldsymbol{\theta}$. In particular, when the confidence C in the investor's experience is large the posterior becomes peaked around the prior value $\boldsymbol{\theta}_0$. On the other hand, when the number of observations T in the time series is large the posterior becomes peaked around the classical estimate $\widehat{\boldsymbol{\theta}}$:

$$f_{\text{po}}(\boldsymbol{\theta}) \begin{array}{l} \nearrow \; \boldsymbol{\theta}_0 \; (C \to \infty) \\ \\ \searrow \; \widehat{\boldsymbol{\theta}} \;\; (T \to \infty). \end{array} \tag{7.4}$$

7.1.2 Summarizing the posterior distribution

The main properties of the posterior distribution are summarized in its location and dispersion parameters, see Figure 7.2.

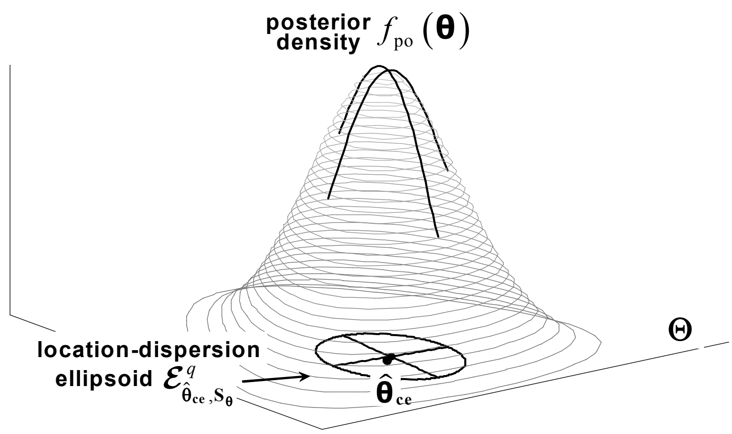

Fig. 7.2. Bayesian posterior distribution and uncertainty set

Location

The location parameter of an S-variate posterior distribution $f_{\text{po}}(\boldsymbol{\theta})$ is an S-dimensional vector $\widehat{\boldsymbol{\theta}}$, see Section 2.4. Since the posterior distribution is determined by the information i_T, in addition to the investor's experience e_C, the location parameter is a *number* that depends on information. This is the definition (7.2) of a classical estimator. Therefore a location parameter of the posterior distribution defines a *classical-equivalent estimator*.

A standard choice for the location parameter of a distribution is its expected value (2.54). Therefore, we introduce the following classical-equivalent estimator of the parameter $\boldsymbol{\theta}$:

$$\begin{aligned}\widehat{\boldsymbol{\theta}}_{\text{ce}}[i_T, e_C] &\equiv \mathrm{E}_{i_T,e_C}\{\boldsymbol{\theta}\} \\ &\equiv \int_{\boldsymbol{\Theta}} \boldsymbol{\theta} f_{\text{po}}(\boldsymbol{\theta}; i_T, e_C)\, d\boldsymbol{\theta}.\end{aligned} \tag{7.5}$$

As it turns out, this classical-equivalent estimator minimizes the estimation error defined by a quadratic loss function as in (4.19). Furthermore, under fairly general conditions, this classical-equivalent estimator is admissible, see Berger (1985).

Another standard choice for the location parameter of a distribution is its mode (2.52). Therefore, we introduce the following classical-equivalent estimator of the parameter $\boldsymbol{\theta}$:

$$\widehat{\boldsymbol{\theta}}_{\text{ce}}\left[i_T, e_C\right] \equiv \operatorname{Mod}_{i_T,e_C}\left\{\boldsymbol{\theta}\right\} \equiv \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\operatorname{argmax}}\, f_{\text{po}}\left(\boldsymbol{\theta}; i_T, e_C\right). \tag{7.6}$$

This classical-equivalent estimator (7.6) based on the mode is equal to the classical-equivalent estimator (7.5) based on the expected value when the posterior is normally distributed, see (2.158). Furthermore, it yields the point of highest concentration of probability in the domain $\boldsymbol{\Theta}$ even when the moments of the posterior distribution are not defined, see Figure 7.2.

A classical-equivalent estimator is an instance of the shrinkage estimators discussed in Section 4.4. For this reason they are called *Bayes-Stein shrinkage estimators*. In the Bayesian context the shrinkage target is represented by the investor's prior experience and the extent of the shrinkage is driven by the relation between the amount of information, i.e. the length T of the time series, and the investor's confidence C in his experience, see Figure 7.1.

In particular, because of (7.4), when the investor's confidence C in his experience is high, the posterior distribution becomes extremely concentrated around the prior $\boldsymbol{\theta}_0$. Therefore the classical-equivalent estimator also shrinks to the point $\boldsymbol{\theta}_0$. Similarly, when the length T of the time series is large, the posterior distribution becomes extremely concentrated around the historical estimate $\widehat{\boldsymbol{\theta}}$. Therefore the classical-equivalent estimator also converges to the historical estimate $\widehat{\boldsymbol{\theta}}$.

Dispersion

A dispersion parameter of the S-variate posterior distribution $f_{\text{po}}(\boldsymbol{\theta})$ is a symmetric and positive $S \times S$ matrix $\mathbf{S}_{\boldsymbol{\theta}}$. Since the posterior is determined by the information i_T and by investor's experience e_C, so is the dispersion parameter.

A standard choice for the dispersion parameter of a distribution is represented by the covariance matrix (2.67), which in this context reads:

$$\mathbf{S}_{\boldsymbol{\theta}}\left[i_T, e_C\right] \equiv \operatorname{Cov}_{i_T,e_C}\left\{\boldsymbol{\theta}\right\} \equiv \int_{\boldsymbol{\Theta}} \left(\boldsymbol{\theta} - \mathrm{E}\left\{\boldsymbol{\theta}\right\}\right)\left(\boldsymbol{\theta} - \mathrm{E}\left\{\boldsymbol{\theta}\right\}\right)' f_{\text{po}}\left(\boldsymbol{\theta}; i_T, e_C\right) d\boldsymbol{\theta}. \tag{7.7}$$

Alternatively, we can consider the modal dispersion (2.65), which in this context reads:

$$\mathbf{S}_{\boldsymbol{\theta}}[i_T, e_C] \equiv \mathrm{MDis}_{i_T, e_C}\{\boldsymbol{\theta}\} \equiv -\left(\left.\frac{\partial^2 \ln f_{\mathrm{po}}(\boldsymbol{\theta}; i_T, e_C)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\mathrm{Mod}\{\boldsymbol{\theta}\}}\right)^{-1}. \tag{7.8}$$

The modal dispersion is equal to the covariance matrix when the posterior is normally distributed:

$$\boldsymbol{\theta} \sim \mathrm{N}\left(\widehat{\boldsymbol{\theta}}_{\mathrm{ce}}, \mathbf{S}_{\boldsymbol{\theta}}\right), \tag{7.9}$$

see (2.159). Furthermore, it provides a measure of the dispersion of the parameter $\boldsymbol{\theta}$ in the range $\boldsymbol{\Theta}$ even when the moments of the posterior are not defined.

Location-dispersion ellipsoid

Together with the classical-equivalent S-dimensional vector $\widehat{\boldsymbol{\theta}}_{\mathrm{ce}}$, the $S \times S$ dispersion matrix $\mathbf{S}_{\boldsymbol{\theta}}$ defines the location-dispersion ellipsoid with radius proportional to q of the estimate of the market parameters $\boldsymbol{\theta}$:

$$\mathcal{E}^q_{\widehat{\boldsymbol{\theta}}_{\mathrm{ce}}, \mathbf{S}_{\boldsymbol{\theta}}} \equiv \left\{\boldsymbol{\theta} \text{ such that } \left(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_{\mathrm{ce}}\right)' \mathbf{S}^{-1}_{\boldsymbol{\theta}} \left(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_{\mathrm{ce}}\right) \leq q^2\right\}, \tag{7.10}$$

see Figure 7.2. Refer to Section 2.4.3 for a thorough discussion of the location-dispersion ellipsoid in a general context.

The location-dispersion ellipsoid defines naturally a self-adjusting uncertainty region for $\boldsymbol{\theta}$. Indeed, we show in Appendix www.7.1 that in the specific case (7.9) where the posterior is normally distributed the following result holds for the probability that the parameters lie within the boundaries of the ellipsoid:

$$\mathbb{P}\left\{\boldsymbol{\theta} \in \mathcal{E}^q_{\widehat{\boldsymbol{\theta}}_{\mathrm{ce}}, \mathbf{S}_{\boldsymbol{\theta}}}\right\} = F_{\chi^2_S}\left(q^2\right), \tag{7.11}$$

where $F_{\chi^2_S}$ is the cumulative distribution function of the chi-square distribution with S degrees of freedom, which is a special case of the gamma cumulative distribution function (1.111). More in general, the least upper bound (2.90) of the Chebyshev inequality applies:

$$\mathbb{P}\left\{\boldsymbol{\theta} \notin \mathcal{E}^q_{\widehat{\boldsymbol{\theta}}_{\mathrm{ce}}, \mathbf{S}_{\boldsymbol{\theta}}}\right\} \leq \frac{S}{q^2}. \tag{7.12}$$

Furthermore, because of (7.4), when the investor's confidence C in his experience is large, the posterior distribution becomes extremely concentrated around the prior $\boldsymbol{\theta}_0$. Therefore the dispersion parameter $\mathbf{S}_{\boldsymbol{\theta}}$ becomes small and the uncertainty ellipsoid (7.10) shrinks to the point $\boldsymbol{\theta}_0$, no matter the radius factor q. Similarly, when the number of observations T in the time series of the market invariants is large, the posterior distribution becomes extremely concentrated around the historical estimate $\widehat{\boldsymbol{\theta}}$. Therefore the dispersion parameter $\mathbf{S}_{\boldsymbol{\theta}}$ becomes small and the uncertainty ellipsoid (7.10) shrinks to the point $\widehat{\boldsymbol{\theta}}$, no matter the radius factor q.

The self-adjusting uncertainty region represented by the location-dispersion ellipsoid (7.10) of the posterior distribution of the parameter $\boldsymbol{\theta}$ plays an important role in robust Bayesian allocation decisions.

7.1.3 Computing the posterior distribution

To compute explicitly the posterior distribution we denote the probability density function of the market invariants by the conditional notation $f(\mathbf{x}|\boldsymbol{\theta})$. In so doing we are implicitly considering the parameters $\boldsymbol{\theta}$ as a random variable, where the true, unknown value $\boldsymbol{\theta}^{\mathrm{t}}$ is the specific instance of this random variable that is chosen by Nature.

Since the market invariants are independent and identically distributed the joint probability density function of the available information (7.1) assuming known the value of the parameters $\boldsymbol{\theta}$ is the product of the probability density functions of the invariants:

$$f_{I_T|\boldsymbol{\theta}}(i_T|\boldsymbol{\theta}) = f(\mathbf{x}_1|\boldsymbol{\theta}) \cdots f(\mathbf{x}_T|\boldsymbol{\theta}), \tag{7.13}$$

see also (4.5).

The investor has some prior knowledge of the parameters, which reflects his experience e_C and is modeled by the prior density $f_{\mathrm{pr}}(\boldsymbol{\theta})$. From the relation between the conditional and the joint probability density functions (2.40) we obtain the expression for the joint distribution of the observations and the market parameters:

$$f_{I_T,\boldsymbol{\theta}}(i_T,\boldsymbol{\theta}) = f_{I_T|\boldsymbol{\theta}}(i_T|\boldsymbol{\theta})\, f_{\mathrm{pr}}(\boldsymbol{\theta}). \tag{7.14}$$

The posterior probability density function is simply the density of the parameters conditional on current information. It follows from the joint density of the observations and the parameters by applying Bayes' rule (2.43), which in this context reads:

$$f_{\mathrm{po}}(\boldsymbol{\theta}; i_T, e_C) \equiv f(\boldsymbol{\theta}|i_T) = \frac{f_{I_T,\boldsymbol{\theta}}(i_T,\boldsymbol{\theta})}{\int_{\boldsymbol{\Theta}} f_{I_T,\boldsymbol{\theta}}(i_T,\boldsymbol{\theta})\, d\boldsymbol{\theta}}. \tag{7.15}$$

By construction, the Bayes posterior distribution smoothly blends the information from the market i_T with the investor's experience e_C, which is modeled by the prior density.

Although the Bayesian approach is conceptually simple, it involves multiple integrations. Therefore, the choices of distributions that allow us to obtain analytical results is quite limited. Parametric models for the investor's prior and the market invariants that give rise to tractable posterior distributions of the market parameters are called *conjugate distributions*.

We present in Sections 7.2 and 7.3 notable conjugate models that allow us to model the markets. If analytical results are not available, one has to resort to numerical simulations, see Geweke (1999).

7.2 Location and dispersion parameters

We present here the Bayesian estimators of the location and the dispersion parameters of the market invariants under the normal hypothesis:

$$\mathbf{X}_t|\boldsymbol{\mu},\boldsymbol{\Sigma} \sim \mathrm{N}\left(\boldsymbol{\mu},\boldsymbol{\Sigma}\right). \tag{7.16}$$

In this setting, the location parameter is the expected value $\boldsymbol{\mu}$ and the scatter parameter is the covariance matrix $\boldsymbol{\Sigma}$. This specification is rich and flexible enough to suitably model real problems, yet the otherwise analytically intractable computations of Bayesian analysis can be worked out completely, see also Aitchison and Dunsmore (1975).

7.2.1 Computing the posterior distribution

The Bayesian estimate of the unknown parameters is represented by the joint posterior distribution of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. In order to compute this distribution we need to collect the information available and to model the investor's experience, i.e. his prior distribution.

Information from the market

The information on the market is contained in the time series (7.1) of the past realizations of the market invariants.

Since we are interested in the estimation of the location parameter $\boldsymbol{\mu}$ and of the scatter parameter $\boldsymbol{\Sigma}$, it turns out sufficient to summarize the historical information on the market into the sample estimator of location (4.98), i.e. the sample mean:

$$\widehat{\boldsymbol{\mu}} \equiv \frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t, \tag{7.17}$$

and the sample estimator of dispersion (4.99), i.e. the sample covariance:

$$\widehat{\boldsymbol{\Sigma}} \equiv \frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{x}_t-\widehat{\boldsymbol{\mu}}\right)\left(\mathbf{x}_t-\widehat{\boldsymbol{\mu}}\right)'. \tag{7.18}$$

Along with the number of observations T in the sample, this is all the information we need from the market. Therefore we can represent this information equivalently as follows:

$$i_T \equiv \left\{\widehat{\boldsymbol{\mu}},\widehat{\boldsymbol{\Sigma}};T\right\}. \tag{7.19}$$

Prior knowledge

We model the investor's prior as a *normal-inverse-Wishart (NIW) distribution*. In other words, it is convenient to factor the joint distribution of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ into the conditional distribution of $\boldsymbol{\mu}$ given $\boldsymbol{\Sigma}$ and the marginal distribution of $\boldsymbol{\Sigma}$.

We model the conditional prior on $\boldsymbol{\mu}$ given $\boldsymbol{\Sigma}$ as a normal distribution with the following parameters:

$$\boldsymbol{\mu}|\boldsymbol{\Sigma} \sim \mathrm{N}\left(\boldsymbol{\mu}_0, \frac{\boldsymbol{\Sigma}}{T_0}\right), \tag{7.20}$$

where $\boldsymbol{\mu}_0$ is an N-dimensional vector and T_0 is a positive scalar.

We model the marginal prior on $\boldsymbol{\Sigma}$ as an inverse-Wishart distribution. In other words, it is easier to model the distribution of the inverse of $\boldsymbol{\Sigma}$, which we assume Wishart-distributed with the following parameters:

$$\boldsymbol{\Sigma}^{-1} \sim \mathrm{W}\left(\nu_0, \frac{\boldsymbol{\Sigma}_0^{-1}}{\nu_0}\right), \tag{7.21}$$

where $\boldsymbol{\Sigma}_0$ is an $N \times N$ symmetric and positive matrix and ν_0 is a positive scalar. For a graphical interpretation of the prior (7.20) and (7.21) in the case $N \equiv 1$ refer to Figure 7.1.

To analyze the role played by the parameters that appear in the above distributions, we first compute the unconditional (marginal) prior on $\boldsymbol{\mu}$. As we show in Appendix www.7.5, this is a multivariate Student t distribution:

$$\boldsymbol{\mu} \sim \mathrm{St}\left(\nu_0, \boldsymbol{\mu}_0, \frac{\boldsymbol{\Sigma}_0}{T_0}\right). \tag{7.22}$$

From this expression we see that the parameter $\boldsymbol{\mu}_0$ in (7.20) reflects the investor's view on the parameter $\boldsymbol{\mu}$. Indeed, from (2.190) we obtain:

$$\mathrm{E}\{\boldsymbol{\mu}\} = \boldsymbol{\mu}_0. \tag{7.23}$$

On the other hand the parameter T_0 in (7.20) reflects his confidence in that view. Indeed, from (2.191) we obtain:

$$\mathrm{Cov}\{\boldsymbol{\mu}\} = \frac{\nu_0}{\nu_0 - 2} \frac{\boldsymbol{\Sigma}_0}{T_0}. \tag{7.24}$$

Therefore a large T_0 corresponds to little uncertainty about the view on $\boldsymbol{\mu}$.

The parameter $\boldsymbol{\Sigma}_0$ in (7.21) reflects the investor's view on the dispersion parameter $\boldsymbol{\Sigma}$. Indeed, from (2.227) we see that the prior expectation reads:

$$\mathrm{E}\left\{\boldsymbol{\Sigma}^{-1}\right\} = \boldsymbol{\Sigma}_0^{-1}. \tag{7.25}$$

On the other hand, the parameter ν_0 in (7.21) describes the investor's confidence in this view. Indeed, from (2.229) we obtain:

$$\mathrm{Cov}\left\{\mathrm{vec}\left[\boldsymbol{\Sigma}^{-1}\right]\right\} = \frac{1}{\nu_0}\left(\mathbf{I}_{N^2} + \mathbf{K}_{NN}\right)\left(\boldsymbol{\Sigma}_0^{-1} \otimes \boldsymbol{\Sigma}_0^{-1}\right), \tag{7.26}$$

where vec is the operator (A.104) that stacks the columns of $\boldsymbol{\Sigma}^{-1}$ into a vector, $\mathbf{I}$ is the identity matrix, $\mathbf{K}$ is the commutation matrix (A.108) and $\otimes$ is the Kronecker product (A.96). Therefore a large value ν_0 corresponds to little uncertainty about the view on $\boldsymbol{\Sigma}^{-1}$ and thus about the view on $\boldsymbol{\Sigma}$.

To summarize, the investor's experience and his confidence are described by the following prior parameters:

$$e_C \equiv \{\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0; T_0, \nu_0\}. \tag{7.27}$$

To determine the specific values of these parameters in financial applications we can use the techniques discussed in Section 7.4.

Posterior distribution

Given the above assumptions on the market, i.e. (7.16), and on the investor's experience, i.e. (7.20) and (7.21), it is possible to carry out the integration in (7.15) explicitly and compute the posterior distribution of the market parameters.

As we show in Appendix www.7.2, the posterior is, like the prior, a normal-inverse-Wishart (NIW) distribution. Indeed, recall (7.17) and (7.18), and define the following additional parameters:

$$T_1\left[i_T, e_C\right] \equiv T_0 + T \tag{7.28}$$

$$\boldsymbol{\mu}_1\left[i_T, e_C\right] \equiv \frac{1}{T_1}\left[T_0\boldsymbol{\mu}_0 + T\widehat{\boldsymbol{\mu}}\right] \tag{7.29}$$

$$\nu_1\left[i_T, e_C\right] \equiv \nu_0 + T \tag{7.30}$$

$$\boldsymbol{\Sigma}_1\left[i_T, e_C\right] \equiv \frac{1}{\nu_1}\left[\nu_0\boldsymbol{\Sigma}_0 + T\widehat{\boldsymbol{\Sigma}} + \frac{\left(\boldsymbol{\mu}_0 - \widehat{\boldsymbol{\mu}}\right)\left(\boldsymbol{\mu}_0 - \widehat{\boldsymbol{\mu}}\right)'}{\frac{1}{T} + \frac{1}{T_0}}\right]. \tag{7.31}$$

Then the posterior distribution of the location parameter conditioned on the dispersion parameter is normal:

$$\boldsymbol{\mu}|\boldsymbol{\Sigma} \sim \mathrm{N}\left(\boldsymbol{\mu}_1, \frac{\boldsymbol{\Sigma}}{T_1}\right); \tag{7.32}$$

and the posterior distribution of the dispersion parameter is inverse-Wishart:

$$\boldsymbol{\Sigma}^{-1} \sim \mathrm{W}\left(\nu_1, \frac{\boldsymbol{\Sigma}_1^{-1}}{\nu_1}\right). \tag{7.33}$$

Also, since both prior and posterior distributions are normal-inverse-Wishart, from (7.22) we immediately derive the unconditional posterior distribution of the location parameter:

$$\boldsymbol{\mu} \sim \mathrm{St}\left(\nu_1, \boldsymbol{\mu}_1, \frac{\boldsymbol{\Sigma}_1}{T_1}\right). \tag{7.34}$$

7.2.2 Summarizing the posterior distribution

We can summarize the main features of the posterior distribution of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ by means of its location-dispersion ellipsoid, as discussed in Section 7.1.2.

We have two options: we can consider the two separate location-dispersion ellipsoids of the marginal posterior distributions of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ respectively, or we can consider the single location-dispersion ellipsoid of the joint posterior distribution of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. Since both approaches find applications in allocation problems, we present both cases.

Marginal posterior distribution of the expected value $\boldsymbol{\mu}$

As far as $\boldsymbol{\mu}$ is concerned, its marginal posterior distribution is the Student t distribution (7.34).

First we compute the classical-equivalent estimator of $\boldsymbol{\mu}$, i.e. a parameter of location of the marginal posterior distribution of $\boldsymbol{\mu}$. Choosing either the expected value (7.5) or the mode (7.6) as location parameter, we obtain from (2.190) the following classical-equivalent estimator:

$$\widehat{\boldsymbol{\mu}}_{\text{ce}}\left[i_T, e_C\right] = \frac{T_0 \boldsymbol{\mu}_0 + T \widehat{\boldsymbol{\mu}}}{T_0 + T}. \tag{7.35}$$

It is easy to check that, as the number of observations T increases, this classical-equivalent estimator shrinks towards the sample mean $\widehat{\boldsymbol{\mu}}$. On the other hand, as the investor's confidence T_0 in his experience regarding $\boldsymbol{\mu}$ increases, the classical-equivalent estimator (7.35) shrinks toward the investor's view $\boldsymbol{\mu}_0$. Notice the symmetric role that the confidence level T_0 and the number of observations T play in (7.35): the confidence level T_0 can be interpreted as the number of "pseudo-observations" that would be necessary in a classical setting to support the investor's confidence about his view $\boldsymbol{\mu}_0$.

Now we turn to the dispersion parameter for $\boldsymbol{\mu}$. Choosing the covariance (7.7) as scatter parameter we obtain from (2.191) the following result:

$$\mathbf{S}_{\boldsymbol{\mu}}\left[i_T, e_C\right] = \frac{1}{T_1} \frac{\nu_1}{\nu_1 - 2} \boldsymbol{\Sigma}_1, \tag{7.36}$$

where the explicit dependence on information and experience is given in (7.28)-(7.31). It can be proved that choosing the modal dispersion (7.8) as scatter parameter the result is simply rescaled by a number close to one.

The location and dispersion parameters (7.35) and (7.36) respectively define the location-dispersion uncertainty ellipsoid (7.10) for $\boldsymbol{\mu}$ with radius proportional to q:

$$\mathcal{E}^q_{\widehat{\boldsymbol{\mu}}_{\text{ce}}, \mathbf{S}_{\boldsymbol{\mu}}} \equiv \left\{ \boldsymbol{\mu} \text{ such that } \left(\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}}_{\text{ce}}\right)' \mathbf{S}_{\boldsymbol{\mu}}^{-1} \left(\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}}_{\text{ce}}\right) \leq q^2 \right\}. \tag{7.37}$$

From (7.36) and the definitions (7.28)-(7.31) we observe that when either the number of observations T or the confidence in the views T_0 tends to infinity, the Bayesian setting becomes the classical setting. Indeed, in this case the uncertainty ellipsoid (7.37) shrinks to the single point $\widehat{\boldsymbol{\mu}}_{\text{ce}}$, no matter the radius factor q. In other words, the marginal posterior distribution of $\boldsymbol{\mu}$ becomes infinitely peaked around its classical-equivalent estimator.

Marginal posterior distribution of the covariance $\boldsymbol{\Sigma}$

As far as $\boldsymbol{\Sigma}$ is concerned, its marginal posterior distribution is the inverse-Wishart distribution (7.33).

First we compute the classical-equivalent estimator of $\boldsymbol{\Sigma}$, i.e. a parameter of location of the marginal posterior distribution of $\boldsymbol{\Sigma}$. Choosing the mode (7.6) as location parameter, we show in Appendix www.7.4 that the ensuing classical-equivalent estimator reads:

$$\widehat{\boldsymbol{\Sigma}}_{\text{ce}}\left[i_T, e_C\right] = \frac{1}{\nu_0 + T + N + 1}\left[\nu_0 \boldsymbol{\Sigma}_0 + T\widehat{\boldsymbol{\Sigma}} + \frac{\left(\boldsymbol{\mu}_0 - \widehat{\boldsymbol{\mu}}\right)\left(\boldsymbol{\mu}_0 - \widehat{\boldsymbol{\mu}}\right)'}{\frac{1}{T} + \frac{1}{T_0}}\right]. \tag{7.38}$$

It can be proved that choosing the expected value (7.5) as location parameter the result is simply rescaled by a number close to one.

It is easy to check that, as the number of observations T increases, the classical-equivalent estimator (7.38) shrinks towards the sample covariance $\widehat{\boldsymbol{\Sigma}}$. On the other hand, as the investor's confidence ν_0 in his experience regarding $\boldsymbol{\Sigma}$ increases, the classical-equivalent estimator (7.38) shrinks toward the investor's view $\boldsymbol{\Sigma}_0$. Notice the symmetric role that the confidence level ν_0 and the number of observations T play in (7.38): the confidence level ν_0 can be interpreted as the number of "pseudo-observations" that would be necessary in a classical setting to support the investor's confidence about his view $\boldsymbol{\Sigma}_0$.

Now we turn to the dispersion parameter for $\boldsymbol{\Sigma}$. Since $\boldsymbol{\Sigma}$ is symmetric, we disregard the redundant elements above the diagonal. In other words we only consider the vector $\operatorname{vech}[\boldsymbol{\Sigma}]$, where vech is the operator that stacks the columns of a matrix skipping the redundant entries above the diagonal. Choosing the modal dispersion (7.8) as scatter parameter, we show in Appendix www.7.4 that the dispersion of $\operatorname{vech}[\boldsymbol{\Sigma}]$ reads:

$$\mathbf{S}_{\boldsymbol{\Sigma}}\left[i_T, e_C\right] = \frac{2\nu_1^2}{\left(\nu_1 + N + 1\right)^3}\left(\mathbf{D}_N'\left(\boldsymbol{\Sigma}_1^{-1} \otimes \boldsymbol{\Sigma}_1^{-1}\right)\mathbf{D}_N\right)^{-1}, \tag{7.39}$$

where $\mathbf{D}_N$ is the duplication matrix (A.113); $\otimes$ is the Kronecker product (A.95); and the explicit dependence on information and experience is given in (7.28)-(7.31). It can be proved that choosing the covariance matrix (7.7) as scatter parameter for $\operatorname{vech}[\boldsymbol{\Sigma}]$, the result is simply rescaled by a number close to one.

The location and dispersion parameters (7.38) and (7.39) respectively define the location-dispersion uncertainty ellipsoid (7.10) for $\mathbf{\Sigma}$ with radius proportional to q:

$$\mathcal{E}^{q}_{\widehat{\mathbf{\Sigma}}_{\mathrm{ce}},\mathbf{S}_{\mathbf{\Sigma}}} \equiv \left\{ \mathbf{\Sigma} : \ \operatorname{vech}\left[\mathbf{\Sigma} - \widehat{\mathbf{\Sigma}}_{\mathrm{ce}}\right]' \mathbf{S}_{\mathbf{\Sigma}}^{-1} \operatorname{vech}\left[\mathbf{\Sigma} - \widehat{\mathbf{\Sigma}}_{\mathrm{ce}}\right] \leq q^2 \right\}. \tag{7.40}$$

Notice that the matrices $\mathbf{\Sigma}$ in this ellipsoid are always symmetric, because the vech operator only spans the non-redundant elements of a matrix. When the radius factor q is small enough, the matrices $\mathbf{\Sigma}$ in this ellipsoid are also positive, because positivity is a continuous property and $\widehat{\mathbf{\Sigma}}_{\mathrm{ce}}$ is positive.

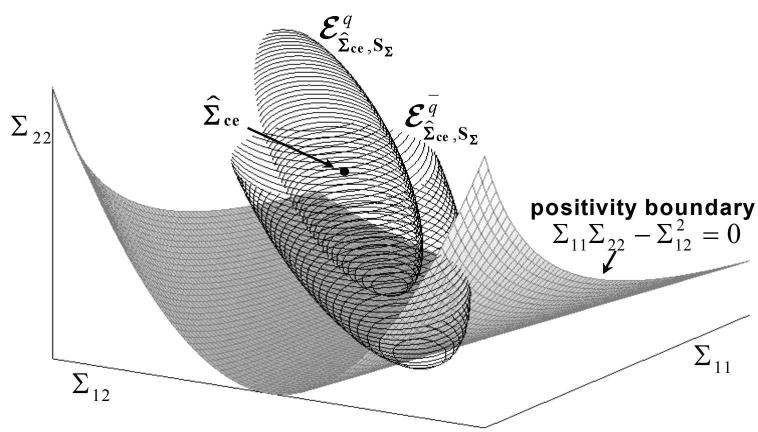

Fig. 7.3. Bayesian location-dispersion ellipsoid for covariance estimation

Consider the case of $N \equiv 2$ market invariants. In this case $\mathbf{\Sigma}$ is a 2×2 matrix:

$$\mathbf{\Sigma} \equiv \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}. \tag{7.41}$$

The symmetry of $\mathbf{\Sigma}$ implies $\Sigma_{12} \equiv \Sigma_{21}$. Therefore a matrix is completely determined by the following three entries:

$$\operatorname{vech}[\mathbf{\Sigma}] = (\Sigma_{11}, \Sigma_{12}, \Sigma_{22})'. \tag{7.42}$$

A symmetric matrix is positive if and only if its eigenvalues are positive. In the 2×2 case, denoting as λ_1 and λ_2 the two eigenvalues, these are positive if and only if the following inequalities are satisfied:

$$\lambda_1 \lambda_2 > 0, \qquad \lambda_1 + \lambda_2 > 0. \tag{7.43}$$

On the other hand, the product of the eigenvalues is the determinant of $\boldsymbol{\Sigma}$ and the sum of the eigenvalues is the trace of $\boldsymbol{\Sigma}$, which are both invariants, see Appendix A.4. Therefore the positivity condition is equivalent to the two conditions below:

$$|\boldsymbol{\Sigma}| \equiv \Sigma_{11}\Sigma_{22} - \Sigma_{12}^2 \geq 0 \tag{7.44}$$

$$\operatorname{tr}(\boldsymbol{\Sigma}) \equiv \Sigma_{11} + \Sigma_{22} \geq 0, \tag{7.45}$$

where the first expression follows from $(A.41)$.

In Figure 7.3 we see that when the radius factor q is small enough, every point of the ellipsoid (7.40) satisfies (7.44)-(7.45). When the radius factor becomes $\overline{q}$, a large enough scalar, the positivity condition is violated.

From (7.39) and the definitions (7.28)-(7.31) we observe that when either the number of observations T or the confidence in the views ν_0 tends to infinity, the Bayesian setting becomes the classical setting. Indeed in this case the uncertainty ellipsoid (7.40) shrinks to the single point $\widehat{\boldsymbol{\Sigma}}_{\text{ce}}$, no matter the radius factor q. In other words, the marginal posterior distribution of $\boldsymbol{\Sigma}$ becomes infinitely peaked around its classical-equivalent estimator.

Joint posterior distribution of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$

We now turn to the analysis of the joint posterior distribution of

$$\boldsymbol{\theta} \equiv \begin{pmatrix} \boldsymbol{\mu} \\ \operatorname{vech}[\boldsymbol{\Omega}] \end{pmatrix}, \tag{7.46}$$

where $\boldsymbol{\Omega} \equiv \boldsymbol{\Sigma}^{-1}$. Indeed, it is much easier to parameterize the joint distribution of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ in terms of the inverse of $\boldsymbol{\Sigma}$.

In Appendix www.7.3 we compute the mode (7.6) of the posterior distribution of $\boldsymbol{\theta}$, which reads:

$$\widehat{\boldsymbol{\theta}}_{\text{ce}}[i_T, e_C] \equiv \begin{pmatrix} \boldsymbol{\mu}_1 \\ \frac{\nu_1 - N}{\nu_1} \operatorname{vech}\left[\boldsymbol{\Sigma}_1^{-1}\right] \end{pmatrix}, \tag{7.47}$$

where the explicit dependence on information and experience is given in (7.28)-(7.31).

In Appendix www.7.3 we also compute the modal dispersion (7.8) of the posterior distribution of $\boldsymbol{\theta}$, which reads:

$$\mathbf{S}_{\boldsymbol{\theta}}[i_T, e_C] = \begin{pmatrix} \mathbf{S}_{\boldsymbol{\mu}} & \mathbf{0}_{N^2 \times (N(N+1)/2)^2} \\ \mathbf{0}_{(N(N+1)/2)^2 \times N^2} & \mathbf{S}_{\boldsymbol{\Omega}} \end{pmatrix}. \tag{7.48}$$

In this expression:

$$\mathbf{S}_{\boldsymbol{\mu}}\left[i_T,e_C\right] \equiv \frac{1}{T_1}\frac{\nu_1}{\nu_1-N}\boldsymbol{\Sigma}_1 \qquad (7.49)$$

$$\mathbf{S}_{\boldsymbol{\Omega}}\left[i_T,e_C\right] \equiv \frac{2}{\nu_1}\frac{\nu_1-N}{\nu_1}\left[\mathbf{D}_N'\left(\boldsymbol{\Sigma}_1\otimes\boldsymbol{\Sigma}_1\right)\mathbf{D}_N\right]^{-1}, \qquad (7.50)$$

where $\mathbf{D}_N$ is the duplication matrix (A.113); $\otimes$ is the Kronecker product (A.95); and the explicit dependence on information and experience is given in (7.28)-(7.31).

The location and dispersion parameters (7.47) and (7.48) define the joint location-dispersion uncertainty ellipsoid (7.10) with radius factor q. It is straightforward to check that all the comments regarding the self-adjusting nature of the location-dispersion ellipsoids (7.37) and (7.40) also apply to the joint location-dispersion ellipsoid.

7.3 Explicit factors

We present here the Bayesian estimators of factor loadings and perturbation dispersion in a factor model under the normal hypothesis for the market. In other words, we consider an affine explicit factor model:

$$\mathbf{X}_t = \mathbf{B}\mathbf{f}_t + \mathbf{U}_t, \qquad (7.51)$$

where the factors $\mathbf{f}_t$ are known and the perturbations, conditioned on the factors, are normally distributed:

$$\mathbf{X}_t|\mathbf{f}_t,\mathbf{B},\boldsymbol{\Sigma} \sim \mathrm{N}\left(\mathbf{B}\mathbf{f}_t,\boldsymbol{\Sigma}\right). \qquad (7.52)$$

In this setting, the parameters to be determined are the factor loadings $\mathbf{B}$ and the dispersion matrix $\boldsymbol{\Sigma}$. This specification is rich and flexible enough to suitably model real problems, yet the otherwise analytically intractable computations of Bayesian analysis can be worked out completely, see also Press (1982).

7.3.1 Computing the posterior distribution

The Bayesian estimate of the unknown parameters is represented by the joint posterior distribution of $\mathbf{B}$ and $\boldsymbol{\Sigma}$. In order to compute this distribution we need to collect the available information from the market and to model the investor's experience, i.e. his prior distribution.

Information from the market

The information on the market is contained in the time series of the past joint realizations of the market invariants and the factors:

$$i_T \equiv \{\mathbf{x}_1, \mathbf{f}_1, \mathbf{x}_2, \mathbf{f}_2, \ldots, \mathbf{x}_T, \mathbf{f}_T\}. \tag{7.53}$$

Since we are interested in the estimation of the factor loadings $\mathbf{B}$ and the scatter parameter $\boldsymbol{\Sigma}$, it turns out sufficient to summarize the historical information on the market into the ordinary least squares estimator (4.126) of the factor loadings, which we report here:

$$\widehat{\mathbf{B}} \equiv \widehat{\boldsymbol{\Sigma}}_{XF} \widehat{\boldsymbol{\Sigma}}_F^{-1}, \tag{7.54}$$

where

$$\widehat{\boldsymbol{\Sigma}}_{XF} \equiv \frac{1}{T} \sum_{t=1}^{T} \mathbf{x}_t \mathbf{f}_t', \quad \widehat{\boldsymbol{\Sigma}}_F \equiv \frac{1}{T} \sum_{t=1}^{T} \mathbf{f}_t \mathbf{f}_t'; \tag{7.55}$$

and the sample covariance of the residuals (4.128), which we report here:

$$\widehat{\boldsymbol{\Sigma}} \equiv \frac{1}{T} \sum_{t=1}^{T} \left(\mathbf{x}_t - \widehat{\mathbf{B}} \mathbf{f}_t \right) \left(\mathbf{x}_t - \widehat{\mathbf{B}} \mathbf{f}_t \right)'. \tag{7.56}$$

Along with the number of observations T in the sample, this is all the information we need from the market. Therefore we can represent the information on the market equivalently in terms of the following parameters:

$$i_T \equiv \left\{ \widehat{\mathbf{B}}, \widehat{\boldsymbol{\Sigma}}; T \right\}. \tag{7.57}$$

Prior knowledge

We model the investor's prior as a *normal-inverse-Wishart (NIW) distribution*. In other words, it is convenient to factor the joint distribution of $\mathbf{B}$ and $\boldsymbol{\Sigma}$ into the conditional distribution of $\mathbf{B}$ given $\boldsymbol{\Sigma}$ and the marginal distribution of $\boldsymbol{\Sigma}$.

We model the conditional prior on $\mathbf{B}$ given $\boldsymbol{\Sigma}$ as a matrix-valued normal distribution (2.181) with the following parameters:

$$\mathbf{B} | \boldsymbol{\Sigma} \sim \mathrm{N}\left(\mathbf{B}_0, \frac{\boldsymbol{\Sigma}}{T_0}, \boldsymbol{\Sigma}_{F,0}^{-1} \right), \tag{7.58}$$

where $\mathbf{B}_0$ is an $N \times K$ matrix, $\boldsymbol{\Sigma}_{F,0}$ is a $K \times K$ symmetric and positive matrix and T_0 is a positive scalar.

We model the marginal prior on $\boldsymbol{\Sigma}$ as an inverse-Wishart distribution. In other words, it is easier to model the distribution of the inverse of $\boldsymbol{\Sigma}$, which we assume Wishart-distributed with the following parameters:

$$\boldsymbol{\Sigma}^{-1} \sim \mathrm{W}\left(\nu_0, \frac{\boldsymbol{\Sigma}_0^{-1}}{\nu_0} \right), \tag{7.59}$$

where $\boldsymbol{\Sigma}_0$ is an $N \times N$ positive definite matrix and ν_0 is a positive scalar.

To analyze the role played by the parameters that appear in the above expressions, we compute the unconditional (marginal) prior on $\mathbf{B}$. We show in Appendix www.7.8 that this distribution is a matrix-valued Student t distribution (2.198) with the following parameters:

$$\mathbf{B} \sim \mathrm{St}\left(\nu_0 + K - N, \mathbf{B}_0, \frac{\nu_0}{\nu_0 + K - N}\mathbf{\Sigma}_0, \frac{\mathbf{\Sigma}_{F,0}^{-1}}{T_0}\right). \tag{7.60}$$

From this expression we see that the parameter $\mathbf{B}_0$ in (7.58) reflects the investor's view on the parameter $\mathbf{B}$. Indeed, from (2.203) we obtain:

$$\mathrm{E}\{\mathbf{B}\} = \mathbf{B}_0. \tag{7.61}$$

On the other hand from (2.206) the parameter $\mathbf{\Sigma}_{F,0}^{-1}$ in (7.58) yields the covariance structure between the m-th and n-th row of $\mathbf{B}$, i.e. the sensitivities of the m-th and n-th market invariant to the factors:

$$\mathrm{Cov}\left\{\mathbf{B}_{(m)}, \mathbf{B}_{(n)}\right\} = \frac{1}{T_0}\frac{\nu_0}{\nu_0 - N + K - 2}\left[\mathbf{\Sigma}_0\right]_{mn}\mathbf{\Sigma}_{F,0}^{-1}. \tag{7.62}$$

This also shows that the parameter T_0 in (7.58) reflects the investor's confidence on his view on $\mathbf{B}$, as a large T_0 corresponds to small variances and covariances in the prior on the factor loadings.

The parameter $\mathbf{\Sigma}_0$ in (7.59) reflects the investor's view on the dispersion parameter $\mathbf{\Sigma}$. Indeed, from (2.227) the prior expectation reads:

$$\mathrm{E}\left\{\mathbf{\Sigma}^{-1}\right\} = \mathbf{\Sigma}_0^{-1}. \tag{7.63}$$

On the other hand, the parameter ν_0 in (7.59) describes the investor's confidence in this view. Indeed, from (2.229) we obtain:

$$\mathrm{Cov}\left\{\mathrm{vec}\left[\mathbf{\Sigma}^{-1}\right]\right\} = \frac{1}{\nu_0}\left(\mathbf{I}_{N^2} + \mathbf{K}_{NN}\right)\left(\mathbf{\Sigma}_0^{-1} \otimes \mathbf{\Sigma}_0^{-1}\right), \tag{7.64}$$

where vec is the operator (A.104) that stacks the columns of $\mathbf{\Sigma}^{-1}$ into a vector, $\mathbf{I}$ is the identity matrix, $\mathbf{K}$ is the commutation matrix (A.108) and $\otimes$ is the Kronecker product (A.96). Therefore a large value ν_0 corresponds to little uncertainty about the view on $\mathbf{\Sigma}^{-1}$ and thus about the view on $\mathbf{\Sigma}$.

To summarize, the investor's experience and his confidence are described by the following prior parameters:

$$e_C \equiv \left\{\mathbf{B}_0, \mathbf{\Sigma}_0, \mathbf{\Sigma}_{F,0}; T_0, \nu_0\right\}. \tag{7.65}$$

To determine the values of these parameters in financial applications we can use the techniques discussed in Section 7.4.

Posterior distribution

Given the above assumptions it is possible to carry out the integration in (7.15) explicitly.

As we show in Appendix www.7.6, the posterior distribution of $\mathbf{B}$ and $\boldsymbol{\Sigma}$ is, like the prior, a normal-inverse-Wishart (NIW) distribution. Indeed, recall (7.54)-(7.56) and define the following additional parameters:

$$T_1 [i_T, e_C] \equiv T_0 + T \tag{7.66}$$

$$\boldsymbol{\Sigma}_{F,1} [i_T, e_C] \equiv \frac{T_0 \boldsymbol{\Sigma}_{F,0} + T \widehat{\boldsymbol{\Sigma}}_F}{T_0 + T} \tag{7.67}$$

$$\mathbf{B}_1 [i_T, e_C] \equiv \left(T_0 \mathbf{B}_0 \boldsymbol{\Sigma}_{F,0} + T \widehat{\mathbf{B}} \widehat{\boldsymbol{\Sigma}}_F \right) \left(T_0 \boldsymbol{\Sigma}_{F,0} + T \widehat{\boldsymbol{\Sigma}}_F \right)^{-1} \tag{7.68}$$

$$\nu_1 [i_T, e_C] \equiv T + \nu_0 \tag{7.69}$$

$$\begin{aligned} \boldsymbol{\Sigma}_1 [i_T, e_C] \equiv \frac{1}{\nu_1} \Big[& T \widehat{\boldsymbol{\Sigma}} + \nu_0 \boldsymbol{\Sigma}_0 + T_0 \mathbf{B}_0 \boldsymbol{\Sigma}_{F,0} \mathbf{B}_0' \\ & + T \widehat{\mathbf{B}} \widehat{\boldsymbol{\Sigma}}_F \widehat{\mathbf{B}}' - T_1 \mathbf{B}_1 \boldsymbol{\Sigma}_{F,1} \mathbf{B}_1' \Big] . \end{aligned} \tag{7.70}$$

Then the dispersion parameter is inverse-Wishart-distributed:

$$\boldsymbol{\Sigma}^{-1} \sim \mathrm{W} \left(\nu_1, \frac{\boldsymbol{\Sigma}_1^{-1}}{\nu_1} \right) . \tag{7.71}$$

On the other hand the distribution of the factor loadings conditioned on the dispersion parameter is a matrix-valued normal distribution (2.181) with the following parameters:

$$\mathbf{B} | \boldsymbol{\Sigma} \sim \mathrm{N} \left(\mathbf{B}_1, \frac{\boldsymbol{\Sigma}}{T_1}, \boldsymbol{\Sigma}_{F,1}^{-1} \right) . \tag{7.72}$$

Also, since prior and posterior are both normal-inverse-Wishart distributions, from (7.60) we immediately derive the unconditional distribution of the factor loadings, which is a matrix-valued Student t distribution:

$$\mathbf{B} \sim \mathrm{St} \left(\nu_1 + K - N, \mathbf{B}_1, \frac{\nu_1}{\nu_1 + K - N} \boldsymbol{\Sigma}_1, \frac{\boldsymbol{\Sigma}_{F,1}^{-1}}{T_1} \right) . \tag{7.73}$$

7.3.2 Summarizing the posterior distribution

We can summarize the main features of the posterior distribution of $\mathbf{B}$ and $\boldsymbol{\Sigma}$ by means of its location-dispersion ellipsoid, as discussed in Section 7.1.2.

We have two options: we can consider the two separate location-dispersion ellipsoids of the marginal posterior distributions of $\mathbf{B}$ and $\boldsymbol{\Sigma}$ respectively, or we can consider the single location-dispersion ellipsoid of the joint distribution of $\mathbf{B}$ and $\boldsymbol{\Sigma}$. Since both approaches find applications in allocation problems, we present both cases.

Marginal posterior distribution of the factor loadings B

As far as $\mathbf{B}$ is concerned, its marginal posterior distribution is the matrix-valued Student t distribution (7.73).

First we compute the classical-equivalent estimator of $\mathbf{B}$, i.e. a parameter of location of the marginal posterior distribution of $\mathbf{B}$. Choosing the expected value (7.5) as location parameter, we obtain from (2.203) the following classical-equivalent estimator of the factor loadings:

$$\widehat{\mathbf{B}}_{\mathrm{ce}}\left[i_T, e_C\right] = \left(T_0 \mathbf{B}_0 \mathbf{\Sigma}_{F,0} + T \widehat{\mathbf{B}} \widehat{\mathbf{\Sigma}}_F\right) \left(T_0 \mathbf{\Sigma}_{F,0} + T \widehat{\mathbf{\Sigma}}_F\right)^{-1}. \tag{7.74}$$

It is easy to check that, as the number of observations T increases, this classical-equivalent estimator shrinks towards the OLS estimator $\widehat{\mathbf{B}}$. On the other hand, as the investor's confidence T_0 in his experience regarding $\mathbf{B}$ increases, the classical-equivalent estimator (7.74) shrinks toward the investor's view $\mathbf{B}_0$. Notice the symmetric role that the confidence level T_0 and the number of observations T play in (7.74): the confidence level T_0 can be interpreted as the number of "pseudo-observations" that would be necessary in a classical setting to support the investor's confidence about his view $\mathbf{B}_0$.

Now we turn to the dispersion parameter for $\mathbf{B}$. Choosing the covariance (7.7) as scatter parameter we obtain from (2.204) the following result:

$$\mathbf{S}_{\mathbf{B}}\left[i_T, e_C\right] = \frac{1}{T_1} \frac{\nu_1}{\nu_1 + K - N - 2} \mathbf{\Sigma}_{F,1}^{-1} \otimes \mathbf{\Sigma}_1, \tag{7.75}$$

where $\otimes$ is the Kronecker product (A.95) and where the explicit dependence on information and experience is given in (7.66)-(7.70).

The location and dispersion parameters (7.74) and (7.75) respectively define the location-dispersion uncertainty ellipsoid (7.10) for $\mathbf{B}$ with radius proportional to q:

$$\mathcal{E}^q_{\widehat{\mathbf{B}}_{\mathrm{ce}}, \mathbf{S}_{\mathbf{B}}} \equiv \left\{ \mathbf{B} : \ \mathrm{vec}\left[\mathbf{B} - \widehat{\mathbf{B}}_{\mathrm{ce}}\right]' \mathbf{S}_{\mathbf{B}}^{-1} \mathrm{vec}\left[\mathbf{B} - \widehat{\mathbf{B}}_{\mathrm{ce}}\right] \leq q^2 \right\}, \tag{7.76}$$

where vec is the operator (A.104) that stacks the columns of a matrix into a vector.

From (7.75) and the definitions (7.66)-(7.70) we observe that when either the number of observations T or the confidence in the views T_0 tends to infinity, the Bayesian setting becomes the classical setting. Indeed, in this case the uncertainty ellipsoid (7.76) shrinks to the single point $\widehat{\mathbf{B}}_{\mathrm{ce}}$, no matter the radius factor q. In other words, the marginal posterior distribution of $\mathbf{B}$ becomes infinitely peaked around its classical-equivalent estimator.

Marginal posterior distribution of the perturbation covariance Σ

As far as $\mathbf{\Sigma}$ is concerned, its marginal posterior distribution is the inverse-Wishart distribution (7.71).

First we compute the classical-equivalent estimator of $\mathbf{\Sigma}$, i.e. a parameter of location of the marginal posterior distribution of $\mathbf{\Sigma}$. Choosing the mode (7.6) as location parameter, we show in Appendix www.7.4 that the ensuing classical-equivalent estimator reads:

$$\widehat{\mathbf{\Sigma}}_{\text{ce}}[i_T, e_C] = \frac{1}{\nu_0 + T + N + 1}\Big[T\widehat{\mathbf{\Sigma}} + \nu_0\mathbf{\Sigma}_0 + T_0\mathbf{B}_0\mathbf{\Sigma}_{F,0}\mathbf{B}_0' + T\widehat{\mathbf{B}}\widehat{\mathbf{\Sigma}}_F\widehat{\mathbf{B}}' - T_1\mathbf{B}_1\mathbf{\Sigma}_{F,1}\mathbf{B}_1'\Big]. \quad (7.77)$$

It is easy to check that, as the number of observations T increases, the classical-equivalent estimator (7.77) shrinks towards the sample covariance $\widehat{\mathbf{\Sigma}}$. On the other hand, as the investor's confidence ν_0 in his experience regarding $\mathbf{\Sigma}$ increases, the classical-equivalent estimator (7.77) shrinks toward the investor's view $\mathbf{\Sigma}_0$. Notice the symmetric role that the confidence level ν_0 and the number of observations T play in (7.77): the confidence level ν_0 can be interpreted as the number of "pseudo-observations" that would be necessary in a classical setting to support the investor's confidence about his view $\mathbf{\Sigma}_0$.

Now we turn to the dispersion parameter for $\mathbf{\Sigma}$. Since $\mathbf{\Sigma}$ is symmetric, we disregard the redundant elements above the diagonal. In other words we only consider the vector vech$[\mathbf{\Sigma}]$, where vech is the operator that stacks the columns of a matrix skipping the redundant entries above the diagonal. Choosing the modal dispersion (7.8) as scatter parameter, we show in Appendix www.7.4 that the dispersion of vech$[\mathbf{\Sigma}]$ reads:

$$\mathbf{S}_{\mathbf{\Sigma}}[i_T, e_C] = \frac{2\nu_1^2}{(\nu_1 + N + 1)^3}\left(\mathbf{D}_N'\left(\mathbf{\Sigma}_1^{-1} \otimes \mathbf{\Sigma}_1^{-1}\right)\mathbf{D}_N\right)^{-1}, \quad (7.78)$$

where $\mathbf{D}_N$ is the duplication matrix $(A.113)$; $\otimes$ is the Kronecker product $(A.95)$; and the explicit dependence on information and experience is given in (7.66)-(7.70).

The location and dispersion parameters (7.77) and (7.78) define the location-dispersion uncertainty ellipsoid (7.10) for $\mathbf{\Sigma}$ of radius proportional to q:

$$\mathcal{E}^q_{\widehat{\mathbf{\Sigma}}_{\text{ce}}, \mathbf{S}_{\mathbf{\Sigma}}} \equiv \left\{\mathbf{\Sigma}: \ \text{vech}\left[\mathbf{\Sigma} - \widehat{\mathbf{\Sigma}}_{\text{ce}}\right]' \mathbf{S}_{\mathbf{\Sigma}}^{-1}\, \text{vech}\left[\mathbf{\Sigma} - \widehat{\mathbf{\Sigma}}_{\text{ce}}\right] \leq q^2\right\}. \quad (7.79)$$

Notice that the matrices $\mathbf{\Sigma}$ in this ellipsoid are always symmetric, because the vech operator only spans the non-redundant elements of a matrix. When the radius factor q is small enough, the matrices $\mathbf{\Sigma}$ in this ellipsoid are also positive, because positivity is a continuous property and $\widehat{\mathbf{\Sigma}}_{\text{ce}}$ is positive, see Figure 7.3.

From (7.78) and the definitions (7.66)-(7.70) we observe that when either the number of observations T or the confidence in the views ν_0 tends to infinity, the Bayesian setting becomes the classical setting. Indeed in this case the uncertainty ellipsoid (7.79) shrinks to the single point $\widehat{\mathbf{\Sigma}}_{\text{ce}}$, no matter the radius factor q. In other words, the marginal posterior distribution of $\mathbf{\Sigma}$ becomes infinitely peaked around its classical-equivalent estimator.

Joint posterior distribution of B and Σ

We now turn to the analysis of the joint posterior distribution of

$$\boldsymbol{\theta} \equiv \begin{pmatrix} \text{vec}\,[\mathbf{B}] \\ \text{vech}\,[\mathbf{\Omega}] \end{pmatrix}, \tag{7.80}$$

where $\mathbf{\Omega} \equiv \mathbf{\Sigma}^{-1}$. Indeed, it is much easier to parameterize the joint distribution of $\mathbf{B}$ and $\mathbf{\Sigma}$ in terms of the inverse of $\mathbf{\Sigma}$.

In Appendix www.7.7 we compute the mode (7.6) of the posterior distribution of $\boldsymbol{\theta}$, which reads:

$$\widehat{\boldsymbol{\theta}}_{\text{ce}}\,[i_T, e_C] \equiv \begin{pmatrix} \mathbf{B}_1 \\ \frac{\nu_1+K-N-1}{\nu_1}\,\text{vech}\left[\mathbf{\Sigma}_1^{-1}\right] \end{pmatrix}, \tag{7.81}$$

where the explicit dependence on information and experience is given in (7.66)-(7.70).

In Appendix www.7.7 we also compute the modal dispersion (7.8) of the posterior distribution of $\boldsymbol{\theta}$, which reads:

$$\mathbf{S}_{\boldsymbol{\theta}}\,[i_T, e_C] = \begin{pmatrix} \mathbf{S}_{\mathbf{B}} & \mathbf{0}_{(NK)^2\times(N(N+1)/2)^2} \\ \mathbf{0}_{(N(N+1)/2)^2\times(NK)^2} & \mathbf{S}_{\mathbf{\Omega}} \end{pmatrix}. \tag{7.82}$$

In this expression:

$$\mathbf{S}_{\mathbf{B}}\,[i_T, e_C] \equiv \frac{1}{T_1}\frac{\nu_1}{\nu_1+K-N-1}\mathbf{K}_{NK}\left(\mathbf{\Sigma}_1 \otimes \mathbf{\Sigma}_{F,1}^{-1}\right)\mathbf{K}_{KN} \tag{7.83}$$

$$\mathbf{S}_{\mathbf{\Omega}}\,[i_T, e_C] \equiv \frac{2}{\nu_1}\frac{\nu_1+K-N-1}{\nu_1}\left[\mathbf{D}_N'\left(\mathbf{\Sigma}_1 \otimes \mathbf{\Sigma}_1\right)\mathbf{D}_N\right]^{-1}, \tag{7.84}$$

where $\mathbf{K}_{NK}$ is the commutation matrix (A.108); $\mathbf{D}_N$ is the duplication matrix (A.113); $\otimes$ is the Kronecker product (A.95); and the explicit dependence on information and experience is given in (7.66)-(7.70).

The location and dispersion parameters (7.81) and (7.82) respectively define the joint location-dispersion uncertainty ellipsoid (7.10) with radius factor q. It is straightforward to check that all the comments regarding the self-adjusting nature of the location-dispersion ellipsoids (7.76) and (7.79) also apply to the joint location-dispersion ellipsoid.

7.4 Determining the prior

In Section 7.1 we discussed how the Bayesian approach to parameter estimation relies on the investor's prior knowledge of the market parameters $\boldsymbol{\theta}$, which is modeled in terms of the prior probability density function $f_{\text{pr}}\,(\boldsymbol{\theta})$.

The parametric expression of the prior density is typically determined by a location parameter $\boldsymbol{\theta}_0$, which corresponds to the "peak" of the prior beliefs, and a set of scalars that define the level of dispersion of the prior density, i.e. the confidence in the prior beliefs.

The confidence in the investor's beliefs is usually left as a free parameter that can be tweaked on a case-by-base basis. Therefore specifying the prior corresponds to determining the value of the location parameter $\boldsymbol{\theta}_0$.

For example, assume that the market consists of equity-like securities. Therefore, the linear returns are market invariants:

$$\mathbf{L}_t \equiv \operatorname{diag}\left(\mathbf{P}_{t-\tau}\right)^{-1} \mathbf{P}_t - \mathbf{1}, \tag{7.85}$$

see Section 3.1.1. Assume that the linear returns are normally distributed:

$$\mathbf{L}_t | \boldsymbol{\mu}, \boldsymbol{\Sigma} \sim \mathrm{N}\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right). \tag{7.86}$$

This is the multivariate normal Bayesian model (7.16), where the prior is determined by the following parameters:

$$\boldsymbol{\theta}_0 \equiv \left(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0\right), \tag{7.87}$$

see (7.23) and (7.25).

In this section we present some techniques to quantify the investor's experience, i.e. to define the prior parameters $\boldsymbol{\theta}_0$ that determine the prior and thus the whole Bayesian estimation process.

These techniques rely on the unconstrained *allocation function*, which is the unconstrained optimal allocation (6.33) considered as a function of the parameters $\boldsymbol{\theta}$ that determine the distribution of the underlying market invariants:

$$\boldsymbol{\theta} \mapsto \boldsymbol{\alpha}\left(\boldsymbol{\theta}\right) \equiv \underset{\boldsymbol{\alpha}}{\operatorname{argmax}} \left\{\mathcal{S}_{\boldsymbol{\theta}}\left(\boldsymbol{\alpha}\right)\right\}. \tag{7.88}$$

To illustrate, we consider the leading example in Section 6.1.

From (7.86) the prices at the investment horizon are normally distributed:

$$\mathbf{P}_{T+\tau}^{\boldsymbol{\mu},\boldsymbol{\Sigma}} \sim \mathrm{N}\left(\boldsymbol{\xi}, \boldsymbol{\Phi}\right), \tag{7.89}$$

where the parameters $\boldsymbol{\xi}$ and $\boldsymbol{\Phi}$ follow from (7.85) and read:

$$\boldsymbol{\xi} \equiv \operatorname{diag}\left(\mathbf{p}_T\right)\left(\mathbf{1} + \boldsymbol{\mu}\right), \quad \boldsymbol{\Phi} \equiv \operatorname{diag}\left(\mathbf{p}_T\right) \boldsymbol{\Sigma} \operatorname{diag}\left(\mathbf{p}_T\right). \tag{7.90}$$

The lower-case notation $\mathbf{p}_T$ in the above expressions stresses that the current prices are realized random variables, i.e. they are known.

The index of satisfaction is the certainty-equivalent (6.21), which after substituting (7.90) reads:

$$\mathrm{CE}_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(\boldsymbol{\alpha}) = \boldsymbol{\alpha}' \operatorname{diag}(\mathbf{p}_T)(\mathbf{1}+\boldsymbol{\mu}) - \frac{1}{2\zeta}\boldsymbol{\alpha}' \operatorname{diag}(\mathbf{p}_T)\boldsymbol{\Sigma}\operatorname{diag}(\mathbf{p}_T)\boldsymbol{\alpha}. \quad (7.91)$$

Maximizing this expression, from the first-order conditions we obtain the allocation function:

$$(\boldsymbol{\mu},\boldsymbol{\Sigma}) \mapsto \boldsymbol{\alpha}(\boldsymbol{\mu},\boldsymbol{\Sigma}) = \zeta \operatorname{diag}(\mathbf{p}_T)^{-1}\boldsymbol{\Sigma}^{-1}(\mathbf{1}+\boldsymbol{\mu}). \quad (7.92)$$

7.4.1 Allocation-implied parameters

Here we present in a more general context the approach proposed by Sharpe (1974) and Black and Litterman (1990), see also Grinold (1996) and He and Litterman (2002).

Typically, investors have a vague, qualitative idea of a suitable value for the prior parameters $\boldsymbol{\theta}_0$. Nonetheless, they usually have a very precise idea of what should be considered a suitable portfolio composition $\boldsymbol{\alpha}_0$, which we call the *prior allocation*.

By inverting the allocation function (7.88), we can set the prior parameters $\boldsymbol{\theta}_0$ as the parameters implied by the prior allocation $\boldsymbol{\alpha}_0$:

$$\boldsymbol{\theta}_0 \equiv \boldsymbol{\theta}(\boldsymbol{\alpha}_0). \quad (7.93)$$

In other words, if the market parameters were $\boldsymbol{\theta}_0$, the optimal allocation would be $\boldsymbol{\alpha}_0$: therefore $\boldsymbol{\theta}_0$ is a prior parameter specification consistent with the prior allocation $\boldsymbol{\alpha}_0$.

In general, the dimension of the market parameters, namely the number S of entries in the vector $\boldsymbol{\theta}$, is larger than the dimension of the market, namely the number N of entries in the vector $\boldsymbol{\alpha}$: therefore the function (7.88) cannot be inverted. This problem can be overcome by pinning down some of the parameters by means of some alternative technique.

In our leading example the N-variate allocation function (7.92) is determined by the $S \equiv N(N+3)/2$ free parameters in $(\boldsymbol{\mu},\boldsymbol{\Sigma})$. Fixing a value $\overline{\boldsymbol{\Sigma}}$ for the covariance, for instance by means of a shrinkage estimate (4.160), we obtain the following inverse function:

$$\boldsymbol{\mu}(\boldsymbol{\alpha}) = \frac{1}{\zeta}\overline{\boldsymbol{\Sigma}}\operatorname{diag}(\mathbf{p}_T)\boldsymbol{\alpha} - \mathbf{1}. \quad (7.94)$$

This function yields the *implied expected returns* of an allocation. Thus we can set the prior (7.87) as follows:

$$\boldsymbol{\mu}_0 \equiv \boldsymbol{\mu}(\boldsymbol{\alpha}_0), \quad \boldsymbol{\Sigma}_0 \equiv \overline{\boldsymbol{\Sigma}}. \quad (7.95)$$

More in general, we can impose a set of constraints $\mathcal{C}$ on the allocation function (7.88). Indeed, imposing constraints on portfolios leads to better out-of-sample results, see Frost and Savarino (1988). This way the allocation function results defined as follows:

$$\boldsymbol{\theta} \mapsto \boldsymbol{\alpha}(\boldsymbol{\theta}) \equiv \underset{\boldsymbol{\alpha} \in \mathcal{C}}{\operatorname{argmax}} \left\{ \mathcal{S}_{\boldsymbol{\theta}}(\boldsymbol{\alpha}) \right\}. \tag{7.96}$$

As in (7.93), the implied prior parameters $\boldsymbol{\theta}_0$ are obtained by first inverting this function, possibly fixing some of the parameters with different techniques, and then evaluating the inverse function in the prior allocation $\boldsymbol{\alpha}_0$.

For instance we can assume a budget constraint:

$$\mathcal{C}_1 : \ \boldsymbol{\alpha}' \mathbf{p}_T = w_T. \tag{7.97}$$

Also, we can impose that specific portfolios, i.e. linear combinations of securities, should not exceed given thresholds:

$$\mathcal{C}_2 : \ \underline{\mathbf{g}} \leq \mathbf{G}\boldsymbol{\alpha} \leq \overline{\mathbf{g}}, \tag{7.98}$$

where the $K \times N$ matrix $\mathbf{G}$ determines the specific portfolios and the K-dimensional vectors $\underline{\mathbf{g}}$ and $\overline{\mathbf{g}}$ determine the upper and lower thresholds respectively.

In Appendix www.7.9 we show that by adding the constraints $\mathcal{C}_1$ and $\mathcal{C}_2$ in our leading example the inverse function (7.94) is replaced by the following expression:

$$\boldsymbol{\alpha} \mapsto \boldsymbol{\mu}(\boldsymbol{\alpha}) + [\operatorname{diag}(\mathbf{p}_T)]^{-1} \mathbf{G}' \left(\overline{\boldsymbol{\gamma}} - \underline{\boldsymbol{\gamma}} \right). \tag{7.99}$$

In this expression $\boldsymbol{\mu}(\boldsymbol{\alpha})$ are the expected returns implied by the constraint (7.97), defined implicitly as follows:

$$\boldsymbol{\mu}(\boldsymbol{\alpha}) - \frac{\mathbf{1}'\overline{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\mu}(\boldsymbol{\alpha})}{\mathbf{1}'\overline{\boldsymbol{\Sigma}}^{-1}\mathbf{1}}\mathbf{1} = \frac{1}{\zeta}\left(\overline{\boldsymbol{\Sigma}} \operatorname{diag}(\mathbf{p}_T)\, \boldsymbol{\alpha} - \frac{w_T}{\mathbf{1}'\overline{\boldsymbol{\Sigma}}^{-1}\mathbf{1}}\mathbf{1} \right); \tag{7.100}$$

and $\left(\overline{\boldsymbol{\gamma}}, \underline{\boldsymbol{\gamma}} \right)$ are the Lagrange multipliers relative to the inequality constraints (7.98) which satisfy the Kuhn-Tucker conditions:

$$\overline{\boldsymbol{\gamma}}, \underline{\boldsymbol{\gamma}} \geq \mathbf{0} \tag{7.101}$$

$$\sum_{n=1}^{N} \underline{\gamma}_k G_{kn} \underline{g}_n = \sum_{n=1}^{N} \overline{\gamma}_k G_{kn} \overline{g}_n = 0, \quad k = 1, \ldots, K. \tag{7.102}$$

This is the result of Grinold and Easton (1998), see also Grinold and Kahn (1999).

7.4.2 Likelihood maximization

A different approach to quantify the investor's experience consists in defining the prior parameters $\boldsymbol{\theta}_0$ as a constrained classical maximum likelihood estimate, where the constraint is imposed in terms of the allocation function, see Jagannathan and Ma (2003) for the specific case which we outline in the example below.

Consider the standard maximum likelihood estimator of the market invariants (4.66), which in the Bayesian notation (7.13) of this chapter reads:

$$\begin{aligned} \widehat{\boldsymbol{\theta}} &\equiv \underset{\boldsymbol{\theta} \in \boldsymbol{\Theta}}{\operatorname{argmax}} \, f_{I_T|\boldsymbol{\theta}} \left(i_T | \boldsymbol{\theta} \right) \\ &= \underset{\boldsymbol{\theta} \in \boldsymbol{\Theta}}{\operatorname{argmax}} \left\{ \sum_{t=1}^{T} \ln f \left(\mathbf{x}_t | \boldsymbol{\theta} \right) \right\}, \end{aligned} \tag{7.103}$$

where the terms $\mathbf{x}_t$ represent the observed time series of the market invariants.

Now consider a set $\mathcal{C}$ of investment constraints, see Frost and Savarino (1988). By means of the allocation function $\boldsymbol{\alpha}(\boldsymbol{\theta})$ defined in (7.88) we select a subset in the domain $\boldsymbol{\Theta}$ of possible values for the parameter market parameters:

$$\widetilde{\boldsymbol{\Theta}} \equiv \{ \boldsymbol{\theta} \in \boldsymbol{\Theta} \text{ such that } \boldsymbol{\alpha}(\boldsymbol{\theta}) \in \mathcal{C} \} . \tag{7.104}$$

In our example (7.91), consider an investor who has no risk propensity, i.e. such that $\zeta \to 0$ in his exponential utility function. Assume there exists a budget constraint and a no-short-sale constraint:

$$\mathcal{C}_1 : \ \boldsymbol{\alpha}' \mathbf{p}_T = w_T, \tag{7.105}$$

$$\mathcal{C}_2 : \ \boldsymbol{\alpha} \geq \mathbf{0}. \tag{7.106}$$

In Appendix www.7.9 we show that the constrained allocation function gives rise to the following constraints for the covariance matrix:

$$\widetilde{\boldsymbol{\Theta}} \equiv \left\{ \boldsymbol{\Sigma} \text{ such that } \boldsymbol{\Sigma} \succeq \mathbf{0}, \boldsymbol{\Sigma}^{-1} \mathbf{1} \geq \mathbf{0} \right\}, \tag{7.107}$$

where the notation "$\succeq \mathbf{0}$" stands for "symmetric and positive".

The prior parameters $\boldsymbol{\theta}_0$ can be defined as the maximum likelihood estimate (7.103) of the market parameters constrained to the subset (7.104). In other words, the prior parameters are defined as follows:

$$\begin{aligned} \boldsymbol{\theta}_0 &\equiv \underset{\boldsymbol{\theta} \in \widetilde{\boldsymbol{\Theta}}}{\operatorname{argmax}} \, f_{I_T|\boldsymbol{\theta}} \left(i_T | \boldsymbol{\theta} \right) \\ &= \underset{\boldsymbol{\theta} \in \widetilde{\boldsymbol{\Theta}}}{\operatorname{argmax}} \left\{ \sum_{t=1}^{T} \ln f \left(\mathbf{x}_t | \boldsymbol{\theta} \right) \right\}. \end{aligned} \tag{7.108}$$

From the log-likelihood under the normal hypothesis (7.86) in terms of the inverse of the covariance $\mathbf{\Omega} \equiv \mathbf{\Sigma}^{-1}$ and the constraints (7.107) we obtain:

$$\mathbf{\Omega}_0 = \underset{\substack{\mathbf{\Omega} \succeq \mathbf{0} \\ \mathbf{\Omega 1} \geq \mathbf{0}}}{\operatorname{argmax}} \left\{ \frac{T}{2} \sum_{t=1}^{T} \ln |\mathbf{\Omega}| - \frac{T}{2} \sum_{t=1}^{T} \operatorname{tr} \left[\widehat{\mathbf{\Sigma}} \mathbf{\Omega} \right] \right\}, \tag{7.109}$$

where $\widehat{\mathbf{\Sigma}}$ is the sample covariance (7.18). In turn, this expression defines the prior $\mathbf{\Sigma}_0 \equiv \mathbf{\Omega}_0^{-1}$ in (7.87).