Centre de Recherche en Économie et Statistique
CREST

Série des Documents de Travail

n° 2011-22

Granularity Theory with Application to Finance and Insurance

C. GOURIÉROUX[1]
P. GAGLIARDINI[2]

[1] University of Lugano, CH-6900 Lugano, Switzerland
[2] CREST and University of Toronto, Canada.

Patrick GAGLIARDINI **Christian GOURIEROUX**

GRANULARITY THEORY WITH APPLICATION TO FINANCE AND INSURANCE

The risk analysis in portfolios of credits, or life insurance contracts, is made difficult by the nonlinearities of risk models, the dependencies between the individual risks, and the large size of the portfolios, which can include several thousands of contracts. The granularity principle has been introduced in the Basel II regulation for credit risk to solve these difficulties when computing the reserves. The principle requires three steps. First, the modelling step considers a Risk Factor Model (RFM), which distinguishes the systematic risks from the unsystematic risks. Second, this model is applied to a virtual portfolio of infinite size, leading to the so-called Asymptotic Risk Factor Model (ARFM). This gives in general explicit formulas for the Value-at-Risk and other risk measures, and thus for the required capital. Third, for a portfolio of large but finite size, closed form approximations are derived from an expansion around the ARFM. This provides the granularity adjustment for the required capital. In fact, the granularity principle can be applied to a variety of related problems. It can be applied for instance for efficient estimation in panel factor models with micro- and macro-dynamics, for improving macro-predictions from micro-data, or for pricing derivatives written on large portfolios. The aim of this book is to provide a first overview of granularity theory by following a progressive pedagogical approach.

This state-of-the-art book on granularity theory is ideal for graduate students, researchers and professionals. All will benefit from the emphasis on practical aspects of financial and insurance risk modeling. Doctoral candidates will appreciate the inclusion of mathematical derivations of the deeper results as well as the more advanced questions concerning risk control and credit derivative pricing. By establishing the link between Basel III and Solvency II regulations, the book also addresses the needs of applied researchers employed by financial institutions. A minimal background in statistics and finance is required, but easily completed by the review chapters included in the book.

Patrick GAGLIARDINI is Associate Professor of Econometrics at the Università della Svizzera Italiana and member of the Swiss Finance Institute. His research interests focus on econometrics and financial econometrics, with special emphasis on factor models, method of moments estimation, and applications in asset pricing and credit risk. On these topics he has published papers in international econometrics and finance journals.

Christian GOURIEROUX is Professor at the University of Toronto and Director of the Insurance and Finance Department at CREST in Paris. He is author of several books in Econometrics, Financial Econometrics, and Credit Risk and has published extensively in international journals both in econometrics an finance. He is member of the scientific committees of the Euronext indexes and of the French Financial Market Authority.

Contents

Chapter 1

Introduction

The granularity principle is a methodology to perform asymptotic expansions for panel models with common factor and large cross-sectional size. The panel observations are doubly indexed by individual and time. The granularity principle consists of two steps:

i) First, one analyzes the **Cross-Sectional Asymptotic (CSA)** model corresponding to a (virtual) panel with infinite cross-sectional size $n = \infty$.

ii) Second, the cross-sectional size is assumed large, but finite, and an expansion in $1/n$ is performed around the asymptotic model.

The granularity approach has been first introduced to analyze the risk in large financial portfolios, and in particular to get accurate approximations of the required capital in the framework of the recent Basel 2 regulation [see BCBS (2001), Gordy (2003) and Chapters 6, 7]. The same principle can be used for large portfolio management, or for pricing derivatives written on large sets of risks such as longevity bonds, or derivatives written on an index of Credit Default Swaps, such as the iTraxx (see Chapter 5). This principle can also be used for analyzing the asymptotic behavior of estimators in large panel models, or for obtaining approximate filtering and prediction formulas of the underlying unobservable factor (Chapters 2-5).

The importance of granularity theory is due to a lack of robustness of the standard asymptotic theorems, such as the Law of Large Numbers (LLN) and the Central Limit Theorem (CLT), in presence of a common factor. We briefly review in Section 1.1 the standard asymptotic theorems with their underlying regularity assumptions. In Section 1.2, we modify the standard regularity assumptions by introducing a common unobservable factor and discuss the new asymptotic behavior of the sample mean of the observations. Finally, in Section 1.3, we present the different panel models with common factor to which the granularity theory will be applied.

1.1 The Basic Asymptotic Theorems

In the basic framework, the asymptotic theorems are presented under a simple set of regularity conditions.

Assumption A.1: *The observations Y_i, $i = 1, \ldots, n$, are independent, identically distributed, with finite second-order moments.*

The observations Y_i can be multidimensional, with dimension K, say. The mean (resp. the variance-covariance matrix) of Y_i is a $(K, 1)$ vector denoted by $m = E(Y_i)$ [resp. a (K, K) matrix denoted by $V(Y_i) = \Sigma$]. The components of the mean vector are the expectations of the components of Y_i. The variance-covariance matrix contains the variances of the components of Y_i on the diagonal, and the covariances between pairs of components out of the diagonal.

Then, we have the two following theorems:

Theorem 1.1: Law of Large Numbers (LLN). *Under Assumption A.1, the sample mean $\bar{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i$ converges almost surely to the theoretical mean m, that is, $\bar{Y}_n \stackrel{a.s.}{\to} m$, as $n \to \infty$.*

Theorem 1.2: Central Limit Theorem (CLT). *Under Assumption A.1, the sample mean is asymptotically (multivariate) Gaussian, that is, $\sqrt{n}(\bar{Y}_n -$*

$m) \xrightarrow{d} N(0, \Sigma)$, where $\xrightarrow{d}$ denotes the convergence in distribution as $n \to \infty$.

Thus, the first term in the asymptotic expansion of $\bar{Y}_n$ is deterministic equal to m (by LLN), whereas the second term is stochastic of order $1/\sqrt{n}$ (by CLT).

The LLN and CLT are used in statistics and econometrics to prove the consistency and asymptotic normality of maximum likelihood and moment-type estimators (under standard regularity assumptions). They can also be used to derive core results in economic and finance theory. As an illustration, let us consider n risky assets $i = 1, \ldots, n$, with unitary price at date t and returns $Y_{i,t+1}, i = 1, \ldots, n$ on period $(t, t+1)$. Let us assume that the risky returns satisfy Assumption A.1 with mean m_t and variance Σ_t, and denote $r_{f,t}$ the riskfree return on the same period.

A portfolio including $1/n$ shares of each risky asset has a unitary price at date t and a return on period $(t, t+1)$ equal to the cross-sectional average return $\bar{Y}_{n,t+1} = \frac{1}{n} \sum_{i=1}^{n} Y_{i,t+1}$. By applying the LLN, we see that $\bar{Y}_{n,t+1}$ tends to m_t. Equivalently, in financial terms:

Proposition 1.3: *Under Assumption A.1, the risk is totally eliminated by diversification for a large size portfolio.*

Since the (asymptotic) portfolio is riskfree, we deduce by no-arbitrage that $m_t = r_{f,t}$ (see Review B.2 for the definition of no-arbitrage).

Proposition 1.4: *Under Assumption A.1 and no-arbitrage, the (conditional) expected return of the individual assets is equal to the riskfree rate.*

Thus, in an economy satisfying Assumption A.1, the individual assets cannot generate a conditional expected return strictly larger than the riskfree rate. Equivalently, they necessary pay a zero risk-premium.

1.2 A Lack of Robustness to Cross-Sectional Dependence

The LLN and CLT can be extended to sequences of variables satisfying weaker conditions than Assumption A.1, for instance to stationary time series, or to variables with heterogenous distributions. However, the limit theorems can be strongly modified for other changes in the basic assumptions.

As an illustration, let us assume that the one-dimensional observations are such that:

$$Y_i = F + u_i, \quad i = 1, \dots, n, \tag{1.1}$$

where $F, u_1, \dots, u_n$, are independent variables, $u_1, \dots, u_n$ have a same distribution with zero-mean and variance σ^2, and F is a random variable with mean μ, variance η^2, and probability density function (pdf) g. The variables Y_i, $i = 1, \dots, n$ have identical marginal distributions with mean $E(Y_i) = \mu$, variance $V(Y_i) = \sigma^2 + \eta^2$. However, these variables are dependent due to the common factor F. For instance, the correlation between two observations is $Corr(Y_i, Y_j) = \dfrac{\eta^2}{\eta^2 + \sigma^2}$, for $i \neq j$.

Let us consider the sample mean. We have:

$$\bar{Y}_n = \frac{1}{n} \sum_{i=1}^{n} Y_i = F + \frac{1}{n} \sum_{i=1}^{n} u_i.$$

By applying the LLN to the average of the idiosyncratic terms u_i, we deduce the following asymptotic behaviour:

Proposition 1.5: *Under factor model (1.1), the sample mean tends to the factor value. In particular this limit is stochastic and different from the common mean $E(Y_i) = \mu$.*

In financial terms, the "idiosyncratic risks" $u_i, i = 1, \dots, n$ can be diversified, but not the common risk[1] F. Thus, diversification cannot totally eliminate

[1] Often called systematic or systemic risk in the financial literature.

the risk.

Let us now consider the asymptotic distribution of the sample mean of variables Y_i.

i) If the cross-sectional dimension n is infinite, we have $\lim\limits_{n\to\infty} \bar{Y}_n = F$ and the asymptotic distribution of the sample mean is simply the distribution of F.

ii) If n is large, but finite, we obtain a more accurate approximation of the distribution of this mean by applying the CLT to $u_1, \ldots, u_n$, conditional on factor F. Conditional on factor F, we have approximately:

$$\bar{Y}_n | F \overset{d}{\sim} N(F, \sigma^2/n).$$

Then, we can integrate out the unobservable factor to get the approximate pdf of $\bar{Y}_n$ as:

$$h_n(y) = \int \frac{1}{\sqrt{2\pi\sigma^2/n}} \exp\left[-\frac{n(y-f)^2}{2\sigma^2}\right] g(f)df,$$

that is, a mixture of Gaussian distributions. In the limiting case $n \to \infty$, the Gaussian kernel concentrates at its mean and h_n tends to g, which corresponds to limiting case i).

In this example we have to distinguish between the **cross-sectional asymptotic** (CSA) analysis corresponding to the virtual situation $n = \infty$ and to the limiting distribution $h_\infty(y) = g(y)$, and the **granularity adjustment** (GA), equal to $h_n(y) - g(y)$, which has to be applied when n is large, but finite.

The derivation above relies on the application of the standard LLN and CLT conditional on factor F. When the conditional application of asymptotic theorems is possible, the model is said to be **infinitely granular** or **infinitely fine grained**. The granularity terminology has been first introduced by Gordy [Gordy (2003)]; see also Wilde (2001) and Martin, Wilde (2002).

1.3 Panel Model with Common Factor

We first define the notion of homogenous population before describing the panel models of interest.

i) Homogenous population

The difference between model (1.1) and the i.i.d. Assumption A.1 is the dependence between observations, which is the same across any pair. This leads to the definition of an exchangeable, or homogenous, set of variables.

Definition 1.6: *A set of variables $Y_i, i = 1, \ldots, n$ is* ***exchangeable*** *(or* ***homogenous****), if and only if the distribution of $Y_1, \ldots, Y_n$ is the same as the distribution of $Y_{\sigma(1)}, \ldots, Y_{\sigma(n)}$, for any permutation $\sigma(\cdot)$ of the set of indexes.*

The notion of exchangeability is also valid for a set of individual histories $Y_i = (Y_{i,1}, \ldots, Y_{i,t}, \ldots), i = 1, \ldots, n$, that is, for panel data.

Loosely speaking, the exchangeability condition requires that all the individuals are equivalent. This condition is satisfied for an i.i.d. sequence of variables, but it is also compatible with a specific form of dependence (called **equidependence**) between the variables, as seen in model (1.1).

A representation theorem for an exchangeable (homogenous) set of variables has been first derived by de Finetti (1931) and extended by Hewitt, Savage (1955). We provide below the version of this theorem appropriate for panel models.

Theorem 1.7: Factor representation of an infinite set of exchangeable histories ($n = \infty$)**.** *The infinite set of histories $Y_i = (Y_{i,t}, t \in \mathbb{N}), i = 1, 2, \ldots,$ is exchangeable, if and only if there exists an underlying factor process $F = (F_t, t \in \mathbb{N})$ such that the individual processes $Y_1, \ldots, Y_n$, are i.i.d. conditional on process F, for any $n \in \mathbb{N}$.*

The underlying factor process is generally multidimensional. Theorem 1.7 implies that an homogenous (exchangeable) set of histories is such that the standard LLN and CLT can generally be applied conditional on factor path F, that is, this set is infinitely fine grained.

ii) Homogenous dynamic panel models

In the rest of the book, we generally consider an homogenous set of individual histories with **state space dynamics**. In terms of autoregressive equations of order one [2], these models can be written as:

State equation:

$$F_t = a(F_{t-1}, \eta_t), \tag{1.2}$$

Measurement equations:

$$y_{i,t} = c(y_{i,t-1}, F_t, \varepsilon_{i,t}), \quad i = 1, \ldots, n, \tag{1.3}$$

where η_t, $t \in \mathbb{N}$, and $\varepsilon_{i,t}$, $i = 1, \ldots, n$, $t \in \mathbb{N}$, are i.i.d standard Gaussian vectors. Thus, the dynamics of individual histories $(y_{i,t}), i = 1, \ldots, n$ is defined in two steps. For fixed factor path, there is an individual dynamics, or **microdynamics**, defined by autoregression (1.3). Then, the common dynamics, or **macrodynamics**, of the factor defined by (1.2) will also influence the individual histories.

Functions a and c can be nonlinear, which will induce complicated serial dependence and codependence between the variables. However, model (1.2)-(1.3) is tractable, since the joint process $(F_t, y_{1,t}, \ldots, y_{n,t})$ depends on the past by lagged values of order 1 only, which is the Markov assumption on this joint process.

In economic or financial applications, special cases of model (1.2)-(1.3) are considered. For instance, the "standard approach" of Basel 2 regulation [3]

[2] They can equivalently be written in terms of transition distributions (see Appendix 1.4). We use in the next chapters one, or the other specification.

[3] The Basel 2 regulation allows for a choice between a basic risk analysis, called **standard approach**, and more sophisticated ones, called **advanced approach**. Ceteris paribus, the required capital is higher under the standard approach. In our framework, an advanced approach may consider multiple and dynamic factors, for instance.

suggests a **static model**, that is a model without micro- or macro-dynamics, and with Gaussian factor, such as:

$$\begin{aligned} \textit{State equation:} \quad F_t &= \eta_t, & (1.4) \\ \textit{Measurement equations:} \quad y_{i,t} &= c(F_t, \varepsilon_{i,t}), \quad i = 1, \ldots, n. & (1.5) \end{aligned}$$

The standard **Gaussian linear state space model**, which underlies the implementation of the linear Kalman filter, assumes:

$$\begin{aligned} \textit{State equation:}\ F_t &= \Phi F_{t-1} + \eta_t, & (1.6) \\ \textit{Measurement equations:}\ y_{i,t} &= \alpha + \beta' F_t + \varepsilon_{i,t}, \quad i = 1, \ldots, n. & (1.7) \end{aligned}$$

In this latter model, the whole dynamics passes through the common factor and this dynamics corresponds to a Gaussian Vector AutoRegressive (VAR) model.

1.4 Summary

In an homogenous population, the dynamics of individual histories can always be represented by means of unobservable dynamic factors. When the joint dynamics of $(F_t, y_{i,t}, i = 1, ..., n)$ admits an autoregressive state space representation, the model is easy to simulate, and the cross-sectional asymptotic analysis easy to interpret, as seen in the next chapters.

REFERENCES

Basel Committee on Banking Supervision (2001): "The New Basel Capital Accord", Second Consultative Paper, Bank of International Settlements, January.

de Finetti, B. (1931): "Funzione Caratteristica di un Fenomeno Aleotorio", Atti della R. Accademia dei Lincei, 6, Memorie, Classe di Scienze Fisiche, Mathematiche e Naturali, 4, 251-299.

Gordy, M. (2003): "A Risk-Factor Model Foundation for Ratings-Based Bank Capital Rules", Journal of Financial Intermediation, 12, 199-232.

Gouriéroux, C., and A., Monfort (1996): *Simulation Based Econometric Methods*, Oxford University Press.

Hewitt, E., and L., Savage (1955): "Symmetric Measures on Cartesian Products", Transaction of the American Mathematical Society, 80, 470-501.

Martin, R., and T., Wilde (2002): "Unsystematic Credit Risk", Risk, November, 123-128.

Wilde, T., (2001): "Probing Granularity", Risk, 14, 103-106.

1.5 Appendix: Autoregression and Transition Density

i) One-dimensional continuously-valued process

Let us consider a one-dimensional process (F_t) with continuous distribution, and denote by $H(f_t|f_{t-1}) = P[F_t < f_t|F_{t-1} = f_{t-1}]$ its transition cumulative distribution function (cdf). We have the following Lemma (see also Review A.1):

Lemma 1.8: *The variable* $\eta_t^* = H(F_t|F_{t-1})$ *is independent of* F_{t-1} *and follows a uniform distribution on* $(0,1)$.

Proof: We have

$$\begin{aligned} P[\eta_t^* < u|F_{t-1} = f_{t-1}] &= P[H(F_t|F_{t-1}) < u|F_{t-1} = f_{t-1}] \\ &= P[F_t < H^{-1}(u|f_{t-1})|F_{t-1} = f_{t-1}] \\ &= H[H^{-1}(u|f_{t-1})|f_{t-1}] = u, \quad \forall u \in (0,1), \end{aligned}$$

where $H^{-1}(u|f_{t-1})$ denotes the inverse of $H(f|f_{t-1})$ with respect to f. The result follows, since $G(u) = u$ is the cdf of the uniform distribution on $(0,1)$.

QED

Thus, $\eta_t = \Phi^{-1}(\eta_t^*)$, where Φ is the cdf of a standard Gaussian distribution, is also independent of F_{t-1} and is $N(0,1)$ distributed. We have $\eta_t = \Phi^{-1}(\eta_t^*) = \Phi^{-1}[H(F_t|F_{t-1})]$. We deduce the autoregression:

$$F_t = H^{-1}[\Phi(\eta_t)|F_{t-1}] = a(F_{t-1}, \eta_t), \text{(say)}.$$

This type of result can be extended to multivariate processes, but also to discrete, or qualitative processes [see e.g. Gouriéroux, Monfort (1996), Section 1.4 and the reference therein].

ii) The exchangeable dynamics in terms of distributions

Let us introduce the information available at time $t-1$:

$$J_{t-1} = (F_{t-1}, y_{1,t-1}, \dots, y_{n,t-1}, F_{t-2}, y_{1,t-2}, \dots, y_{n,t-2}, \dots).$$

The conditions equivalent to (1.2)-(1.3) are the following:

State equation: The conditional distribution of F_t given J_{t-1} depends on the past by means of lagged factor value F_{t-1} only. The transition density of the factor is denoted by $g(f_t|f_{t-1})$.

Measurement equations: Conditional on (J_{t-1}, F_t), the variables $y_{1,t}, \dots, y_{n,t}$ are independent. The conditional distribution of $y_{i,t}$ given (J_{t-1}, F_t) depends on $(y_{i,t-1}, F_t)$ only and this dependence is identical for all individuals. The conditional pdf is denoted by $h(y_{i,t}|y_{i,t-1}, f_t)$, with function h independent of individual i.

iii) The joint distribution of individual histories

We deduce the joint density of $(F_t, y_{1,t}, \dots, y_{n,t}, t = 1, \dots, T)$ given the initial values $f_0, y_{1,0}, \dots, y_{n,0}$. It is given by:

$$\prod_{t=1}^{T} \left\{ \left(\prod_{i=1}^{n} h(y_{i,t}|y_{i,t-1}, f_t) \right) g(f_t|f_{t-1}) \right\}.$$

Then, by integrating out the unobservable factor path, we get the joint density of the individual histories only (given J_0) as :

$$\int .. \int \prod_{t=1}^{T} \left\{ \left(\prod_{i=1}^{n} h(y_{i,t}|y_{i,t-1}, f_t) \right) g(f_t|f_{t-1}) \right\} \prod_{t=1}^{T} df_t.$$

This joint density involves an integral of a very large dimension, i.e., a dimension equal to the number of dates multiplied by the number of common factors, which explains the need for tractable approximations of this density.

Chapter 2

Gaussian Static Factor Model

The linear static factor model with Gaussian errors is a benchmark in panel econometrics [see e.g. Rao (1971), Harville (1977)], portfolio management [Markowitz (1952), Lintner (1965)], and arbitrage pricing theory [Ross (1976), (1982), Chamberlain, Rothschild (1983)]. This type of panel model is completely analyzed in this chapter. In Section 2.1, we first discuss the model and its structure. Then we make explicit the granularity adjustment for the estimation of micro- and macro-parameters in Section 2.2. Granularity adjustment for portfolio management is considered in Section 2.3. For both applications, we discuss the introduction of individual heterogeneity in the basic exchangeable model.

2.1 The Model

The panel model considered in this section is known in the literature as the **variance-component**, or **random effect** model. Its simplest version allows for a closed form expression of the maximum likelihood estimators, easy to interpret and to analyze [see e.g. Searle (1971)].

i) The regressions

Let us first introduce the state and measurement equations [see (1.2)-

(1.3)]. We assume one-dimensional observations $y_{i,t}$ and factor F_t, and impose a linear static structure. Then, the state equation is:

$$F_t = u_t, \tag{2.1}$$

whereas the measurement equations are:

$$y_{i,t} = F_t + \varepsilon_{i,t}, \quad i = 1, \ldots, n, \tag{2.2}$$

where (u_t) and $(\varepsilon_{i,t})$ are i.i.d. Gaussian variables, $u_t \sim N(\mu, \eta^2)$ and $\varepsilon_{i,t} \sim N(0, \sigma^2)$. Note that the errors in both the state and measurement equations have not been standardized and that the error in the state equation is not zero-mean.

The model (2.1)-(2.2) is called Gaussian **Linear Single Risk Factor** (LSRF) model. It involves two types of parameters: μ and η^2 are macro-parameters associated with the common factor, whereas σ^2 is a micro-parameter summarizing the individual (or idiosyncratic) risk. We will come back later on the micro- or macro-interpretations of these parameters. The model (2.1)-(2.2) has been used rather early in the literature on risky individual contracts. This is the Buhlmann model considered in the actuarial science, which is the basis for **credibility theory** [Buhlmann (1967), Buhlmann, Straub (1970)].

ii) First- and second-order moments

Let us denote by $\tilde{y}_t = (y_{1,t}, \ldots, y_{n,t})'$ the vector of individual observations at date t. We have:

$$E(\tilde{y}_t) = \mu e, \tag{2.3}$$

where e is the $(n, 1)$ vector with unitary components $e = (1, \ldots, 1)'$. The variance-covariance matrix of $\tilde{y}_t$ is:

$$V(\tilde{y}_t) = \sigma^2 Id + \eta^2 ee' = \Omega, \text{ say}, \tag{2.4}$$

whereas the random vectors $\tilde{y}_t$ and $\tilde{y}_{t'}$ corresponding to two different dates are uncorrelated. The dependence between individual observations (i.e. the cross-dependence) is captured by the term $\eta^2 ee'$ in the variance-covariance matrix, which makes Ω non diagonal.

It is interesting to analyze more deeply the structure of variance-covariance matrix Ω. For this purpose, let us first remark that the matrix $M_1 = ee'/n$ [resp. $M_2 = Id - ee'/n$] is the orthogonal projector on the 1-dimensional linear space generated by vector e (resp. on the $(n-1)$-dimensional linear space orthogonal to the space generated by vector e). We recall that a matrix M is an orthogonal projector if it is symmetric and idempotent, that is, $M' = M$ and $M^2 = M$. The variance-covariance matrix can be decomposed in terms of orthogonal projectors as follows:

$$\Omega = \sigma^2 (Id - \frac{ee'}{n}) + \lambda^2 \frac{ee'}{n}, \tag{2.5}$$

with:

$$\lambda^2 = \sigma^2 + n\eta^2. \tag{2.6}$$

The decomposition (2.5) can be used to derive the **spectral decomposition** [1] and the inverse of matrix Ω (see Appendix 2.4).

Proposition 2.1: i) *The matrix Ω admits as eigenvalues σ^2, with multiplicity order $n-1$, and λ^2 with multiplicity order 1. The eigenspace associated with λ^2 is the space $\mathcal{E}$ generated by vector e. The eigenspace associated with σ^2 is the vector space $\mathcal{E}^\perp$ orthogonal to $\mathcal{E}$. In particular,* $\det \Omega = (\sigma^2)^{n-1}\lambda^2$.

ii) *The inverse of Ω is:*

$$\Omega^{-1} = \frac{1}{\sigma^2}(Id - \frac{ee'}{n}) + \frac{1}{\lambda^2}\frac{ee'}{n}.$$

iii) The cross-sectional distribution of the observations

In terms of the transition and measurements pdf's [see Appendix 1.4 ii)], model (2.1)-(2.2) can be specified as:

$$g(f_t; \mu, \eta^2) = \frac{1}{\sqrt{2\pi\eta^2}} \exp\left\{ -\frac{(f_t - \mu)^2}{2\eta^2} \right\}, \tag{2.7}$$

[1] The spectral decomposition of a matrix is the set of its eigenvalues and the associated eigenvectors.

and

$$h(y_{it}|f_t;\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(y_{i,t}-f_t)^2}{2\sigma^2}\right\}. \tag{2.8}$$

Thus, the density of $\tilde{y}_t$ is [see Appendix 1.4 iii)]:

$$\begin{aligned}
& l(\tilde{y}_t;\sigma^2,\mu,\eta^2) \\
= & \int \prod_{i=1}^{n} h(y_{i,t}|f_t;\sigma^2) g(f_t;\mu,\eta^2) df_t \\
= & \int \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_{i,t}-f_t)^2\right\} \frac{1}{\sqrt{2\pi\eta^2}} \exp\left\{-\frac{1}{2\eta^2}(f_t-\mu)^2\right\} df_t.
\end{aligned} \tag{2.9}$$

This joint pdf has a simplified expression, which can be derived directly by noting that the vector $\tilde{y}_t$ is Gaussian $\tilde{y}_t \sim N(\mu e, \Omega)$ [see Subsection ii)]. We deduce that :

$$l(\tilde{y}_t;\sigma^2,\mu,\eta^2) = \frac{1}{(2\pi)^{n/2}(\det\Omega)^{1/2}} \exp\left\{-\frac{1}{2}(\tilde{y}_t-\mu e)'\Omega^{-1}(\tilde{y}_t-\mu e)\right\}.$$

By Proposition 2.1, we know that:

$$\det\Omega = (\sigma^2)^{n-1}\lambda^2, \qquad \Omega^{-1} = \frac{1}{\sigma^2}(Id - \frac{ee'}{n}) + \frac{1}{\lambda^2}\frac{ee'}{n}.$$

We deduce:

$$\begin{aligned}
& l(\tilde{y}_t;\sigma^2,\mu,\lambda^2) \\
= & \frac{1}{(2\pi)^{n/2}(\sigma^2)^{\frac{n-1}{2}}(\lambda^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}(\tilde{y}_t-\mu e)'(Id-\frac{ee'}{n})(\tilde{y}_t-\mu e)\right. \\
& \left. - \frac{1}{2\lambda^2}(\tilde{y}_t-\mu e)'\frac{ee'}{n}(\tilde{y}_t-\mu e)\right\}.
\end{aligned}$$

This likelihood is written in terms of the new parameter λ^2 and this parameter involves the number n of cross-sectional observations [see equation (2.6)].

Since $(Id - \frac{ee'}{n})e = 0$, we get:

$$\begin{aligned} l(\tilde{y}_t; \sigma^2, \mu, \lambda^2) &= \frac{1}{(2\pi)^{n/2}(\sigma^2)^{\frac{n-1}{2}}(\lambda^2)^{1/2}} \exp\left\{ -\frac{1}{2\sigma^2}\tilde{y}_t'(Id - \frac{ee'}{n})\tilde{y}_t \right. \\ &\left. - \frac{1}{2\lambda^2}\frac{1}{n}[e'(\tilde{y}_t - \mu e)]^2 \right\}. \end{aligned}$$

Let us now introduce the following cross-sectional summary statistics of the panel data:

$$\bar{y}_t = \frac{1}{n}\sum_{i=1}^{n} y_{i,t}, \tag{2.10}$$

is the cross-sectional sample mean of the individual data, and:

$$\sigma_t^2 = \frac{1}{n}\sum_{i=1}^{n}(y_{i,t} - \bar{y}_t)^2, \tag{2.11}$$

is its cross-sectional variance. It is easily checked that:

$$e'(\tilde{y}_t - \mu e) = n(\bar{y}_t - \mu), \tag{2.12}$$

$$\tilde{y}_t'(Id - \frac{ee'}{n})\tilde{y}_t = n\sigma_t^2. \tag{2.13}$$

By substituting in the expression of the pdf, we get:

$$\begin{aligned} & l(\tilde{y}_t; \sigma^2, \mu, \lambda^2) \\ = & \frac{1}{(2\pi)^{n/2}(\sigma^2)^{(n-1)/2}(\lambda^2)^{1/2}} \exp\left\{ -\frac{n}{2\sigma^2}\sigma_t^2 - \frac{n}{2\lambda^2}(\bar{y}_t - \mu)^2 \right\}. \end{aligned} \tag{2.14}$$

This means that the pair $(\bar{y}_t, \sigma_t^2)$ defined in (2.10)-(2.11) is a sufficient statistic to capture all the information contained in the observations of date t.

2.2 Estimation of the Parameters

i) Maximum likelihood (ML) estimators

From the simplified expression (2.14) of the cross-sectional pdf, we deduce the log-likelihood function:

$$\begin{aligned} L_{n,T}(\sigma^2, \mu, \lambda^2) &= \sum_{t=1}^{T} \log l(\tilde{y}_t; \sigma^2, \mu, \lambda^2) \\ &= -\frac{nT}{2}\log(2\pi) - \frac{T(n-1)}{2}\log\sigma^2 - \frac{T}{2}\log\lambda^2 \\ &- \frac{n}{2\sigma^2}\sum_{t=1}^{T}\sigma_t^2 - \frac{n}{2\lambda^2}\sum_{t=1}^{T}(\bar{y}_t - \mu)^2. \end{aligned} \tag{2.15}$$

The log-likelihood function can be first optimized with respect to the mean parameter μ. The first- order condition is:

$$\begin{aligned} & \frac{\partial L_{n,T}}{\partial \mu}(\sigma^2, \mu, \lambda^2) = 0 \\ \iff & \sum_{t=1}^{T}(\bar{y}_t - \mu) = 0 \\ \iff & \mu = \frac{1}{T}\sum_{t=1}^{T}\bar{y}_t. \end{aligned} \tag{2.16}$$

Let us now introduce the following additional summary statistics of the observations:

$$\bar{\bar{y}} = \frac{1}{nT}\sum_{t=1}^{T}\sum_{i=1}^{n} y_{i,t}, \tag{2.17}$$

is the sample average over all observations,

$$B(y) = \frac{1}{T}\sum_{t=1}^{T}(\bar{y}_t - \bar{\bar{y}})^2, \tag{2.18}$$

is the variance **between** the cross-sectional averages of different dates, and:

$$W(y) = \frac{1}{T}\sum_{t=1}^{T}\sigma_t^2, \tag{2.19}$$

is the sample average of the variances **within** dates.

From (2.16), the ML estimator of the mean is:

$$\hat{\mu}_{nT} = \bar{\bar{y}}. \tag{2.20}$$

Then, the log-likelihood can be concentrated with respect to parameter μ. The concentrated log-likelihood, that is the log-likelihood preliminarily optimized with respect to μ, is:

$$\begin{aligned} L^c_{n,T}(\sigma^2, \lambda^2) &= -\frac{nT}{2}\log(2\pi) - \frac{T(n-1)}{2}\log\sigma^2 - \frac{T}{2}\log\lambda^2 \\ &- \frac{nT}{2\sigma^2}W(y) - \frac{nT}{2\lambda^2}B(y). \end{aligned} \tag{2.21}$$

This concentrated log-likelihood is the sum of a function of σ^2 and a function of λ^2. Therefore, the optimizations with respect to these parameters can be performed separately. We get:

$$\hat{\sigma}^2_{n,T} = \frac{n}{n-1}W(y), \tag{2.22}$$

$$\hat{\lambda}^2_{n,T} = nB(y), \tag{2.23}$$

and the ML estimator of η^2 is deduced by using equation (2.6).

The results above are summarized in the following proposition:

Proposition 2.2: *The maximum likelihood estimators of the parameters are:*

$$\hat{\mu}_{n,T} = \bar{\bar{y}}, \qquad \hat{\sigma}^2_{n,T} = \frac{n}{n-1}W(y), \qquad \hat{\eta}^2_{n,T} = B(y) - \frac{W(y)}{n-1}.$$

Thus, the ML estimators of the parameters have closed form expressions in the basic variance component model. They are functions of the total empirical mean, and of the within and between variances. Their properties will be deduced from the properties of these three summary statistics.

ii) Asymptotic behaviour

As usual in panel models, there exist different settings for asymptotic analysis, since we can have either n large, or T large, or both n and T large. The appropriate asymptotic setting depends on the application and the available data. In the applications we are interested in, the individuals are typically financial assets, contracts, or companies, and the number n can be of the order of some thousands. The order of the time dimension T is related to the frequency of observations. The number of observations dates can be about $20 - 50$ with yearly data (e.g., for corporate rating histories), or of the order of some hundreds with monthly data (e.g. for households mortgages).

Let us assume that the time and cross-sectional dimensions T and n are both large, and focus on the effect of n.

(*) The ML estimator of μ can be decomposed as:

$$\hat{\mu}_{n,T} \quad = \quad \frac{1}{nT}\sum_{t=1}^{T}\sum_{i=1}^{n}(F_t + \varepsilon_{it}) = \frac{1}{T}\sum_{t=1}^{T} F_t + \frac{1}{nT}\sum_{t=1}^{T}\sum_{i=1}^{n}\varepsilon_{it}.$$

Thus:

$$\hat{\mu}_{n,T} - \mu = \frac{1}{T}\sum_{t=1}^{T}(F_t - \mu) + \frac{1}{nT}\sum_{t=1}^{T}\sum_{i=1}^{n}\varepsilon_{i,t}. \tag{2.24}$$

The first term in decomposition (2.24) is Gaussian, zero-mean, with order $1/\sqrt{T}$, while the second term is Gaussian, zero-mean, with order $1/\sqrt{nT}$. Moreover, the two terms are independent. We have:

$$V(\hat{\mu}_{n,T}) = \frac{\eta^2}{T} + \frac{\sigma^2}{nT}.$$

When $n = \infty$, only the first term matters, and the speed of convergence of the estimator of μ corresponds to the number of observation dates T, which is compatible with the interpretation of μ as a macro-parameter. When n is large, but finite, the second term in the decomposition provides the necessary adjustment for the cross-sectional effect, that is, the granularity adjustment.

(**) The ML estimator of σ^2 can be written as:

$$\begin{aligned}
\hat{\sigma}^2_{n,T} &= \frac{n}{n-1}\frac{1}{T}\sum_{t=1}^{T}\left[\frac{1}{T}\sum_{i=1}^{n}(y_{i,t}-\bar{y}_t)^2\right] \\
&= \frac{n}{n-1}\frac{1}{T}\sum_{t=1}^{T}\left[\frac{1}{n}\sum_{i=1}^{n}(\varepsilon_{i,t}-\bar{\varepsilon}_t)^2\right] \\
&= \frac{n}{n-1}W(\varepsilon).
\end{aligned}$$

Since the idiosyncratic errors are i.i.d. Gaussian, the variables $\sum_{i=1}^{n}(\varepsilon_{i,t}-\bar{\varepsilon}_t)^2/\sigma^2$, t varying, are independent, with chi-square distribution $\chi^2(n-1)$. Then, by the CLT we deduce that that $\hat{\sigma}^2_{n,T}$ tends to σ^2 at speed $1/\sqrt{nT}$. This speed involves the total number of observations nT, which corresponds to the interpretation of the idiosyncratic variance as a micro-parameter. In the limiting case $n=\infty$, the estimator coincides with the true parameter value σ^2. Thus, the order of the granularity adjustment is equal to the order of $\frac{n}{n-1}W(\varepsilon)-\sigma^2$, that is, to the order $1/\sqrt{nT}$.

(***) Let us finally consider the behavior of the estimator of η^2 for infinite cross-sectional size $n=\infty$. For $n=\infty$, we get:

$$\begin{aligned}
\hat{\eta}^2_{n,T} &= B(y) = \frac{1}{T}\sum_{t=1}^{T}(F_t+\bar{\varepsilon}_t-\bar{F}-\bar{\bar{\varepsilon}})^2 \\
&= \frac{1}{T}\sum_{t=1}^{T}(F_t-\bar{F})^2,
\end{aligned}$$

since $\bar{\varepsilon}_t=\bar{\bar{\varepsilon}}=0$, by the LLN. Thus, the CSA ML estimator is equivalent to the empirical variance of the factor; it tends to η^2 at speed $1/\sqrt{T}$, corresponding to the interpretation of η^2 as a macro-parameter.

The results above are summarized in the Proposition below:

Proposition 2.3: i) *If $n=\infty$, the estimator of σ^2 is constant equal to the unknown true parameter value. The estimators of μ and η^2 are stochastic;*

they tend to the true value of the associated parameters when T tends to infinity, at the macro-speed $1/\sqrt{T}$.

ii) *The different estimators are consistent, if both n and T tend to infinity, with different speeds of adjustment, that are the micro-speed $1/\sqrt{nT}$ for parameter σ^2, the macro-speed $1/\sqrt{T}$ for parameters μ and η^2.*

The difference between cases i) and ii) in Proposition 2.3 provides the granularity adjustments for the distributions of the maximum likelihood estimators.

iii) Finite sample behaviour

Let us now investigate the finite sample behaviour of estimators $\hat{\mu}_{n,T}$, $\hat{\sigma}^2_{n,T}$ and $\hat{\eta}^2_{n,T}$. Figures 2.1, 2.2 and 2.3 display the pdfs of these estimators for different combinations of cross-sectional and time sample sizes, that are $n = T = 20$, $n = 20$ and $T = 100$, $n = 100$ and $T = 20$, $n = T = 100$. The true values of the parameters are $\mu = 0$, $\sigma^2 = 1$, $\eta^2 = 1$. The pdfs of the estimators are obtained by simulating $10,000$ independent samples, computing the estimates of μ, σ^2 and η^2 for each sample, and then computing the kernel density of the estimates.

[Insert Figure 2.1: Pdf of estimator $\hat{\mu}_{n,t}$.]

[Insert Figure 2.2: Pdf of estimator $\hat{\sigma}^2_{n,t}$.]

[Insert Figure 2.3: Pdf of estimator $\hat{\eta}^2_{n,t}$.]

In Figure 2.1 it is seen that the pdf of the estimator of μ is centered around the true value of the parameter. The pdf gets more concentrated when the time dimension T of the sample increases (compare left and right panels), but is rather insensitive to the cross-sectional dimension n (compare upper and lower panels). This finding is compatible with the asymptotic analysis in the previous section and the macro-speed $1/\sqrt{T}$ of parameter μ. The pdf of estimator $\hat{\mu}_{n,T}$ is Gaussian for all sample sizes, since the estimator is a linear transformation of the Gaussian data.

Figure 2.2 shows that the variance of the estimator of σ^2 decreases, when either the time dimension T or the cross-sectional dimension n increase. This confirms the interpretation of σ^2 as a micro-parameter with rate of convergence $1/\sqrt{nT}$. The pdf of $\hat{\sigma}^2_{n,T}$ appears rather close to a Gaussian distribution for the considered sample sizes. In fact, from (2.25) the finite sample distribution of $\hat{\sigma}^2_{n,T}$ is $\sigma^2\chi^2[(n-1)T]/[(n-1)T]$. Finally, in Figure 2.3 it is seen that the distribution of the estimator of parameter η^2 gets more concentrated around the true value when the time dimension increases, but not when the cross-sectional dimension alone increases. Indeed, we have seen in the previous section that parameter η^2 admits a macro interpretation and its estimator features a $1/\sqrt{T}$ rate of convergence. For sample size $T = 20$, the distribution of estimator $\hat{\eta}^2_{n,T}$ is rather far from Gaussian, even for $n = 100$. The distribution is close to Gaussian for $T = 100$.

iv) Choice of the state-space representation

There exist different state space representations of a same dynamic system with unobservable factor since the notion of factor is not defined in a unique way. For instance system (2.1)-(2.2) can be equivalently written as:

System (1):

State equation: $F_t = \mu + \eta u_t, \quad u_t \sim IIN(0,1),$

Measurement equations: $y_{i,t} = F_t + \sigma\varepsilon_{i,t}, \quad \varepsilon_{i,t} \sim IIN(0,1).$

System (2):

State equation: $F_t = \eta u_t, \quad u_t \sim IIN(0,1),$

Measurement equations: $y_{i,t} = \mu + F_t + \sigma\varepsilon_{i,t}, \quad \varepsilon_{i,t} \sim IIN(0,1).$

System (3):

State equation: $F_t = u_t, \quad u_t \sim IIN(0,1),$

Measurement equations: $y_{i,t} = \mu + \eta F_t + \sigma\varepsilon_{i,t}, \quad \varepsilon_{i,t} \sim IIN(0,1).$

For a relevant economic interpretation, it is preferable to select a representation including the micro-parameters in the measurement equations and the macro-parameters in the state equation. We deduce ex-post from the analysis of the asymptotic properties of the estimators (see Proposition 2.3) that the appropriate state-space representation is System (1), that is, the initial representation (2.1)-(2.2).

v) Model with observed heterogeneity

The results derived for exchangeable panel models can be extended to models including **observed heterogeneity**. To highlight this point, let us consider the following extension of model (2.1)-(2.2):

State equation: $F_t = u_t, \quad u_t \sim IIN[\mu, \eta^2]$,

Measurement equations: $y_{i,t} = \beta_i F_t + \varepsilon_{i,t}, \quad \varepsilon_{i,t} \sim IIN(0, \sigma^2)$,
where β_i, $i = 1, \ldots, n$, are known scalars.

The parameter β_i represents the sensitivity of observation $y_{i,t}$ to factor F_t. In the model above the sensitivities can differ across individuals. The sensitivities are usually called beta's in the financial literature, which justifies our notation.

By following an approach similar to the method used for the model without heterogeneity, we get (see Appendix 2.4):

$$\begin{aligned} L_{n,T}(\sigma^2, \mu, \lambda^2) \quad &= \quad -\frac{nT}{2}\log(2\pi) - \frac{T(n-1)}{2}\log\sigma^2 - \frac{T}{2}\log\lambda^2 \\ &\quad -\frac{1}{2\sigma^2}\sum_{t=1}^{T}\tilde{y}_t'(Id - \beta\beta'/\beta'\beta)\tilde{y}_t - \frac{1}{2\lambda^2}\sum_{t=1}^{T}\frac{[\beta'(\tilde{y}_t - \mu\beta)]^2}{\beta'\beta}, \end{aligned}$$

where the new parameter λ^2 is equal to $\lambda^2 = \sigma^2 + \eta^2\beta'\beta$.

Let us focus on the estimation of the factor mean μ. By writing the first-order condition with respect to μ, we get:

$$\hat{\mu}_{n,T}(\beta) = \frac{1}{T}\sum_{t=1}^{T}\frac{\beta'\tilde{y}_t}{\beta'\beta}. \tag{2.25}$$

This estimator admits a two step interpretation. Let us consider the model for given date t, i.e,

$$y_{i,t} = \beta_i F_t + \varepsilon_{i,t}, \quad i = 1, \ldots, n, \quad \varepsilon_{it} \sim IIN(0, \sigma^2).$$

This cross-sectional equation could be considered as a regression model, with unknown regression parameter F_t. In this case, F_t would be approximated by the cross-sectional OLS estimator:

$$\hat{F}_{n,t} = \beta' \tilde{y}_t / \beta' \beta. \tag{2.26}$$

Intuitively, the common factor expectation $\mu = E(F_t)$ is accurately approximated by:

$$\mu \sim \frac{1}{T} \sum_{t=1}^{T} F_t \sim \frac{1}{T} \sum_{t=1}^{T} \hat{F}_{n,t},$$

which is exactly formula (2.25).

Let us now derive the factor decomposition of the estimator $\hat{\mu}_{n,T}(\beta)$. We get:

$$\hat{\mu}_{n,T}(\beta) - \mu = \frac{1}{T} \sum_{t=1}^{T} (F_t - \mu) + \frac{1}{T} \sum_{t=1}^{T} \frac{\beta' \tilde{\varepsilon}_t}{\beta' \beta},$$

where $\tilde{\varepsilon}_t = (\varepsilon_{1,t}, \ldots, \varepsilon_{n,t})'$. The finite sample distribution of this difference is Gaussian with zero mean and variance:

$$\frac{\eta^2}{T} + \frac{\sigma^2}{T} \frac{1}{\sum_{i=1}^{n} \beta_i^2} = \frac{\eta^2}{T} + \frac{\sigma^2}{nT} \frac{1}{(\bar{\beta}_n)^2 + \sigma^2_{\beta,n}},$$

where $\bar{\beta}_n = \frac{1}{n} \sum_{i=1}^{n} \beta_i, \sigma^2_{\beta,n} = \frac{1}{n} \sum_{i=1}^{n} (\beta_i - \bar{\beta})^2$ are the empirical mean and variance of the sensitivity coefficients, respectively.

Let us assume that the individual heterogeneity is well-distributed across individuals in the sense that $\bar{\beta}_\infty = \lim_{n \to \infty} \bar{\beta}_n$ and $\sigma^2_{\beta,\infty} = \lim_{n \to \infty} \sigma^2_{\beta,n}$ exist. It is always possible to assume $\bar{\beta}_\infty = 1$, possibly by changing the definition of parameters μ, η^2. Thus, for n large, the distribution is approximately Gaussian with variance $\frac{\eta^2}{T} + \frac{\sigma^2}{nT} \frac{1}{1+\sigma^2_{\beta,\infty}}$. We get the following Proposition:

Proposition 2.4: *In a Gaussian static factor model with observed beta heterogeneity, the distribution of the ML estimator of μ for $n = \infty$ does not depend on the individual heterogeneity and has a variance proportional to $1/T$. The granularity adjustment, i.e. the term of order $1/(nT)$ in the variance, depends on the individual heterogeneity [2] by means of the variance of the sensitivity coefficients.*

The estimator is the least accurate, when $\sigma^2_{\beta,\infty} = 0$, that is, when the distribution of the betas is the most concentrated.

vi) Granularity Adjustment for factor prediction

Let us still consider the static factor model with observed heterogeneity. The theoretical prediction of F_t given all observations $y_{i,t}, i = 1, \ldots, n, t = 1, \ldots, T$, i.e. the **smoothed value** of F_t, is:

$$\begin{aligned} E[F_t|y] &= E(F_t|\tilde{y}_t) \\ &= E(F_t) + Cov\ (F_t, \tilde{y}_t) V(\tilde{y}_t)^{-1} (\tilde{y}_t - E(\tilde{y}_t)), \end{aligned}$$

by using standard results for Gaussian random vectors (see Review A.5). We have:

$$\begin{aligned} E(F_t) &= \mu, E(\tilde{y}_t) = \mu\beta, \\ Cov(F_t, \tilde{y}_t) &= Cov(F_t, \beta F_t) = \eta^2 \beta', \\ V(\tilde{y}_t) &= \Omega = \sigma^2 Id + \eta^2 \beta\beta'. \end{aligned}$$

By using the results in Appendix 2.4, we get:

$$E(F_t|y) = \mu + \eta^2 \beta' \left[\frac{1}{\sigma^2} (Id - \frac{\beta\beta'}{\beta'\beta}) + \frac{1}{\lambda^2} \frac{\beta\beta'}{\beta'\beta} \right] (\tilde{y}_t - \mu\beta),$$

[2] Also called **concentration** [Lutkebohmert (2008)].

that is,

$$
\begin{aligned}
E(F_t|y) &= \mu + \frac{\eta^2\beta'\beta}{\lambda^2}(\hat{F}_{nt}-\mu) \\
&= \hat{F}_{nt} - \frac{\sigma^2}{\sigma^2+\eta^2\beta'\beta}(\hat{F}_{nt}-\mu) \\
&\sim \hat{F}_{n,t} - \frac{\sigma^2}{\sigma^2+n\eta^2[1+\sigma^2_{\beta,\infty}]}(\hat{F}_{n,t}-\mu) \\
&\sim \hat{F}_{n,t} - \frac{\sigma^2}{n\eta^2[1+\sigma^2_{\beta,\infty}]}(\hat{F}_{n,t}-\mu), \qquad (2.27)
\end{aligned}
$$

when n is large. We deduce that:

(*) The cross-sectional OLS estimator $\hat{F}_{n,t}$ of F_t is an accurate approximation of the smoothed factor value if $n=\infty$. In other words, $\hat{F}_{n,t}$ is the CSA optimal predictor of F_t.

(**) The cross-sectional OLS estimator has to be corrected for large, but finite, sample size n. The granularity adjustment for prediction is equivalent to:

$$
\frac{\hat{\sigma}^2_{nT}}{n\hat{\eta}^2_{n,T}\left(1+\sigma^2_{\beta,n}\right)}\left(\hat{F}_{nt}-\frac{1}{T}\sum_{t=1}^{T}\hat{F}_{n,t}\right),
$$

after substitution of the parameters by consistent estimates.

2.3 Mean-Variance Portfolio Management

In this section, we consider a static linear factor model for excess asset returns and analyze the standard mean-variance portfolio management. This allows to distinguish the effects of the common factor and idiosyncratic errors on the efficient allocation and Sharpe performance, respectively.

We assume that the excess asset returns on period $(t-1,t)$, that are the differences between the risky and riskfree returns, satisfy the model with heterogeneity of Section 2.2 iv), namely:

$$
y_{i,t} = \beta_i F_t + \varepsilon_{i,t}, \quad i=1,\ldots,n,
$$

where $F_t \sim IIN(\mu, \eta^2)$ and $\varepsilon_{i,t} \sim IIN(0, \sigma^2)$. Thus, the expected excess returns are $E(y_{i,t}) = \beta_i \mu$, and the idiosyncratic risk is measured by $V(\varepsilon_{i,t}) = \sigma^2$. There exists a systematic source of risk, through the common factor. This creates an additional individual risk of $\beta_i^2 \eta^2$, but also a dependence between excess returns of two different risky assets, since the correlations:

$$corr\ (y_{i,t}, y_{j,t}) = \frac{\beta_i \beta_j \eta^2}{(\beta_i^2 \eta^2 + \sigma^2)^{1/2} (\beta_j^2 \eta^2 + \sigma^2)^{1/2}},$$

are non-zero for $i \neq j$.

i) The mean-variance efficient allocation

Let us consider a mean-variance efficient allocation based on the n risky assets and the riskfree asset, held at time t for horizon 1. The vector of efficient allocations in the n risky assets is proportional to [3] [Markowitz (1952), and Review B.1]:

$$a_{n,t} = V_t(\tilde{y}_{t+1})^{-1} E_t(\tilde{y}_{t+1}),$$

where E_t and V_t denote the conditional expectation and variance, respectively, given the information at date t. Due to the static assumption, the conditional and unconditional moments coincide and the efficient allocation is time independent, given by:

$$\begin{aligned} a_n &= V(\tilde{y}_{t+1})^{-1} E(\tilde{y}_{t+1}) \\ &= \Omega^{-1} \mu \beta \\ &= \left[\frac{1}{\sigma^2} (Id - \frac{\beta\beta'}{\beta'\beta}) + \frac{1}{\lambda^2} \frac{\beta\beta'}{\beta'\beta} \right] \mu\beta, \end{aligned}$$

that is,

$$a_n = \frac{\mu}{\sigma^2 + \eta^2 \beta'\beta} \beta. \qquad (2.28)$$

The associated Sharpe performance, that is, the marginal expected return

[3] with a scale depending on the absolute risk aversion of the investor.

adjusted for risk of the n risky assets [Sharpe (1966) and Review B.1], is:

$$\begin{aligned} S_n &= E(\tilde{y}_{t+1})'V(\tilde{y}_{t+1})^{-1}E(\tilde{y}_{t+1}) \\ &= \mu^2\beta'\left[\frac{1}{\sigma^2}(Id - \frac{\beta\beta'}{\beta'\beta}) + \frac{1}{\lambda^2}\frac{\beta\beta'}{\beta'\beta}\right]\beta \\ &= \frac{\mu^2}{\lambda^2}\beta'\beta \\ &= \frac{\mu^2\beta'\beta}{\sigma^2+\eta^2\beta'\beta}. \end{aligned} \tag{2.29}$$

We get the next result.

Proposition 2.5: i) *The efficient allocation in the LSRF model with heterogeneity is:*

$$a_n = \frac{\mu}{\sigma^2+\eta^2\beta'\beta}\beta.$$

ii) *The associated Sharpe performance is:*

$$S_n = \frac{\mu^2\beta'\beta}{\sigma^2+\eta^2\beta'\beta} = \frac{\mu^2}{\eta^2} - \frac{\sigma^2}{\eta^2}\frac{\mu^2}{\sigma^2+\eta^2\beta'\beta}.$$

As usual in such a factor model, the vector of efficient allocations is proportional to the vector of beta's. The Sharpe performance depends on the beta's by means of $\beta'\beta$ and is an increasing function of this quantity.

ii) Large portfolio

Proposition 2.5 provides the explicit expressions of the efficient allocation and Sharpe performance. Let us now study their behaviours for large portfolio size, that is, for large n. Let us recall that:

$$\bar{\beta}_\infty = 1, \text{ and } \beta'\beta \sim n(1+\sigma^2_{\beta,\infty}).$$

We deduce the following Corollary:

Corollary 2.6: *We have:* $\lim_{n\to\infty} a_{n,j} = 0,$ *for any portfolio component* j; $S_\infty = \lim_{n\to\infty} S_n = \mu^2/\eta^2$.

To understand the result above, let us consider the excess return of the whole portfolio. Indeed, even if the allocation in each single asset tends to zero, the whole risky portfolio return does not necessarily vanishes, due to the increase in the number n of included assets. More precisely, we have:

$$\begin{aligned} a_n'\tilde{y}_t &= \frac{\mu\beta'}{\sigma^2+\eta^2\beta'\beta}(\beta F_t + \tilde{\varepsilon}_t) \\ &= \frac{\mu\beta'\beta}{\sigma^2+\eta^2\beta'\beta}F_t + \frac{\mu\beta'\tilde{\varepsilon}_t}{\sigma^2+\eta^2\beta'\beta}. \end{aligned} \tag{2.30}$$

Since $\beta'\tilde{\varepsilon}_t \sim N[0, \sigma^2\beta'\beta]$ is of order $\sqrt{n}$, we deduce that:

$$a_n'\tilde{y}_t \sim \frac{\mu}{\eta^2}F_t, \tag{2.31}$$

does not vanish asymptotically. The results are summarized below.

Proposition 2.7: *For an infinitely large portfolio, the efficient risky allocation is constructed to perfectly hedge the common factor. In particular, the Sharpe performance of the n assets tends to the Sharpe performance of the common factor, namely* $\dfrac{(EF_t)^2}{V(F_t)} = \dfrac{\mu^2}{\eta^2} = S_\infty$.

From a financial point of view, the common factor does not correspond a priori to the return of a tradable asset. Nevertheless, the efficient portfolio a_n defines a new asset, which is tradable, and mimicks perfectly factor F_t when $n = \infty$. It is called the asymptotic **mimicking portfolio**. This portfolio diversifies the idiosyncratic risks to capture the relevant common risk.

iii) Granularity adjustment

In practice the set of assets available to an investor is large, but not "asymptotically large". The portfolio performance is therefore influenced by a residual of undiversified idiosyncratic risk. To account for this residual risk, we can consider the next terms in the expansion with respect to n of the Sharpe performance. For this purpose, let us assume that the square of the beta's are also well-diversified across individuals in the sense that:

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\beta_i^2 - (1+\sigma^2_{\beta,\infty})\right] = \Delta_n \xrightarrow{d} N(0,\Delta), \text{ say.}$$

We have:

$$
\begin{aligned}
S_n &= \frac{\mu^2}{\eta^2}\left(1+\frac{\sigma^2}{\eta^2\beta'\beta}\right)^{-1} \\
&= \frac{\mu^2}{\eta^2}\left[1-\frac{\sigma^2}{\eta^2\beta'\beta}+O(1/n^2)\right] \\
&= \frac{\mu^2}{\eta^2}\left[1-\frac{\sigma^2}{\eta^2}\frac{1}{n(1+\sigma^2_{\beta,\infty})}\frac{1}{1+\dfrac{\Delta_n}{\sqrt{n}(1+\sigma^2_{\beta,\infty})}}+O(1/n^2)\right] \\
&= \frac{\mu^2}{\eta^2}-\frac{1}{n}\frac{\mu^2\sigma^2}{\eta^4}\frac{1}{1+\sigma^2_{\beta,\infty}}+\frac{1}{n\sqrt{n}}\frac{\mu^2\sigma^2}{\eta^4}\frac{\Delta_n}{(1+\sigma^2_{\beta,\infty})^2}+O(1/n^2).
\end{aligned}
$$

We deduce the Proposition below:

Proposition 2.8: i) *The second term (or granularity adjustment) in the expansion of the Sharpe performance is deterministic, of order* $1/n$. *It involves the Sharpe performance of the factor, the ratio of the idiosyncratic and factor risks, and a measure of heterogeneity (concentration).*

ii) *The third term in the expansion is of order* $1/(n\sqrt{n})$, *and is stochastic. It captures the uncertainty of the squared beta's distribution.*

Such expansions can also be performed for the efficient allocation, or for the whole net portfolio return [see Gouriéroux, Monfort (2011), where the extension to portfolio management under short-sell restrictions is also considered].

2.4 Summary

The Gaussian linear single risk factor model is often used since it is simple to understand and to implement. In particular we get closed form expressions for the maximum likelihood estimators, for the predictions of the latent factor, for the mean-variance efficient allocation and the associated Sharpe performance. These closed form expressions can be used to disentangle the CSA and granularity adjustment components of the object of interest.

R E F E R E N C E S

Buhlmann, H. (1967): “Experience Rating and Credibility”, ASTIN Bulletin, 4, 199-207.

Buhlmann, H., and L., Straub (1970): “Credibility for Loss Ratios”, ARCH, 1972-2.

Chamberlain, G., and M., Rothschild (1983): "Arbitrage, Factor Structure and Mean-Variance Analysis in Large Asset Markets”, Econometrica, 51, 1281-1304.

Gourieroux, C., and A., Monfort (2011): "Granularity Adjustment for Efficient Portfolio”, forthcoming Econometric Reviews.

Harville, D. (1977): “ML Approaches to Variance Components Estimation and Related Problems”, JASA, 72, 320-340.

Lintner, J. (1965): "The Valuation of Risky Assets and the Selection of Risky Investments in Stock Portfolio and Capital Budgets”, Review of Economic and Statistics, 47, 13-37.

Lutkebohmert, E. (2008): "Concentration Risk in Credit Portfolio”, Springer Verlag.

Markowitz, H. (1952): "Portfolio Selection”, The Journal of Finance, 7, 77-91.

Rao, C. (1971): “Estimation of Variance and Covariance Components : MINQUE Theory”, J. Multiv. Analysis, 1, 257-275.

Ross, S. (1976): "The Arbitrage Theory of Capital Asset Pricing”, Journal of Economic Theory, 13, 341-360.

Ross, S. (1982): "On the General Validity of the Mean-Variance Approach in Large Markets”, in Financial Economics : Essays in Honor of Paul Cootner, ed., by W., Sharpe and C., Cootner, Prentice Hall.

Searle, S. (1971): *Linear Models*, Wiley.

Sharpe, W. (1966): "Mutual Fund Performance", Journal of Business, 36, 119-138.

2.5 Appendix: Structure of the Variance-Covariance Matrix

Let us consider a variance-covariance matrix of the type:

$$\Omega = \sigma^2 Id + \eta^2 \beta\beta',$$

where σ^2, η^2 are two positive scalars and β a vector of dimension n. By introducing the orthogonal projectors $\beta\beta'/\beta'\beta$ and $Id - \beta\beta'/\beta'\beta$, we can write:

$$\Omega = \sigma^2(Id - \beta\beta'/\beta'\beta) + (\sigma^2 + \eta^2\beta'\beta)(\beta\beta'/\beta'\beta). \quad \text{(a.1)}$$

This equation provides the spectral decomposition of matrix Ω. Its eigenvalues are:

σ^2, with multiplicity order $n-1$, and associated eigenspace the orthogonal of the space generated by vector β;

$\lambda^2 = \sigma^2 + \eta^2\beta'\beta$, with multiplicity order 1, and eigenspace the space generated by β.

In particular:

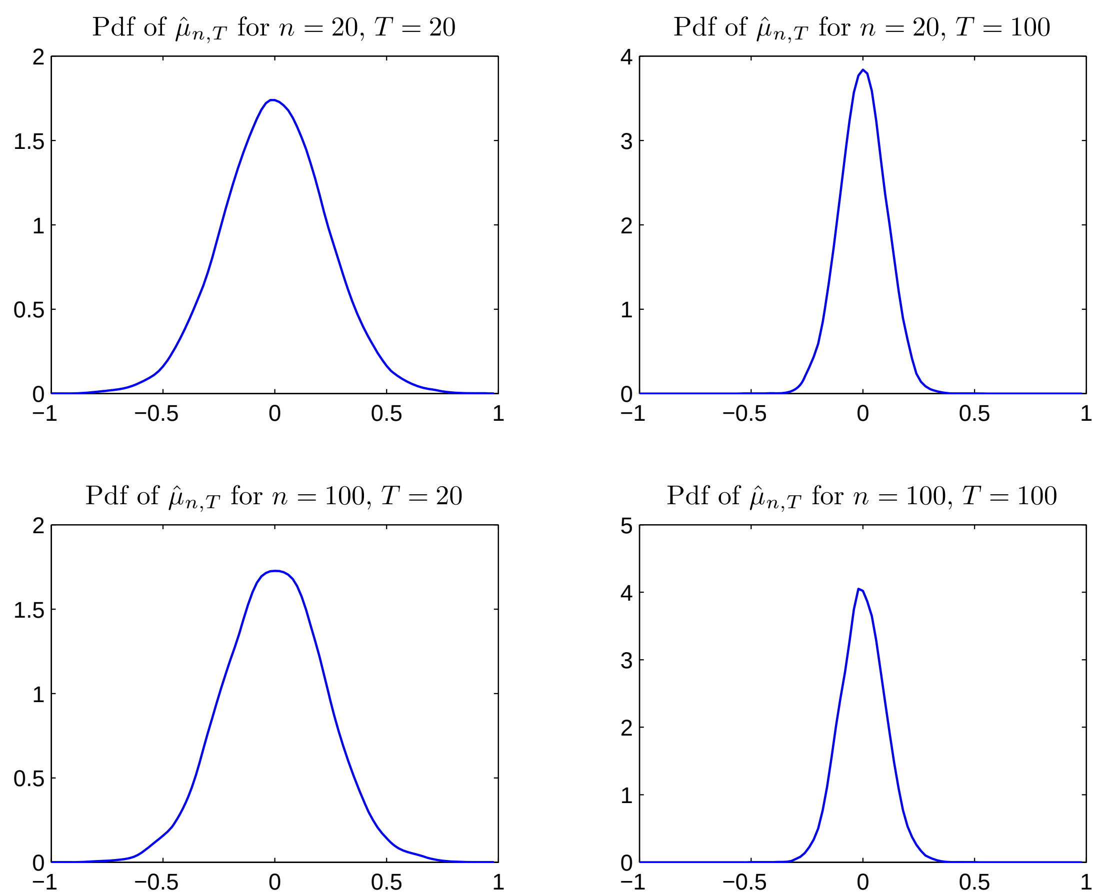
$$\det \Omega = (\sigma^2)^{n-1}\lambda^2,$$

since it is equal to the product of the eigenvalues taking into account their multiplicity orders, and:

$$\Omega^{-1} = \frac{1}{\sigma^2}(Id - \beta\beta'/\beta'\beta) + \frac{1}{\lambda^2}(\beta\beta'/\beta'\beta),$$

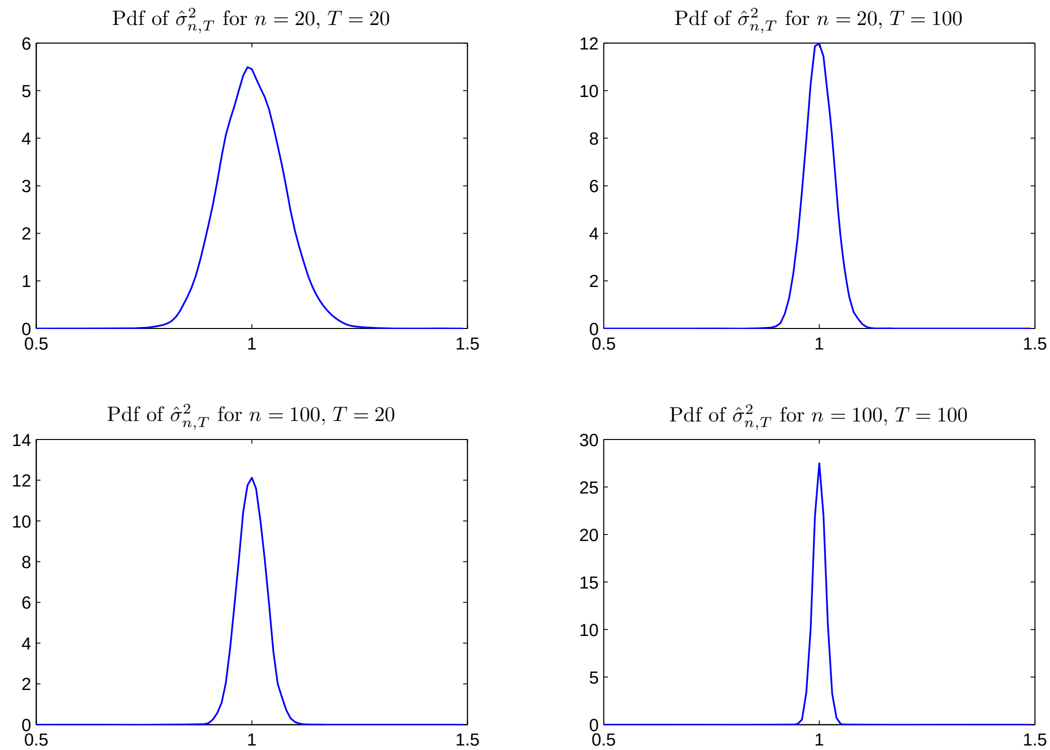
as easily checked by computing the product of this latter matrix with matrix Ω.

Figure 2.1: Pdf of estimator $\hat{\mu}_{n,t}$.



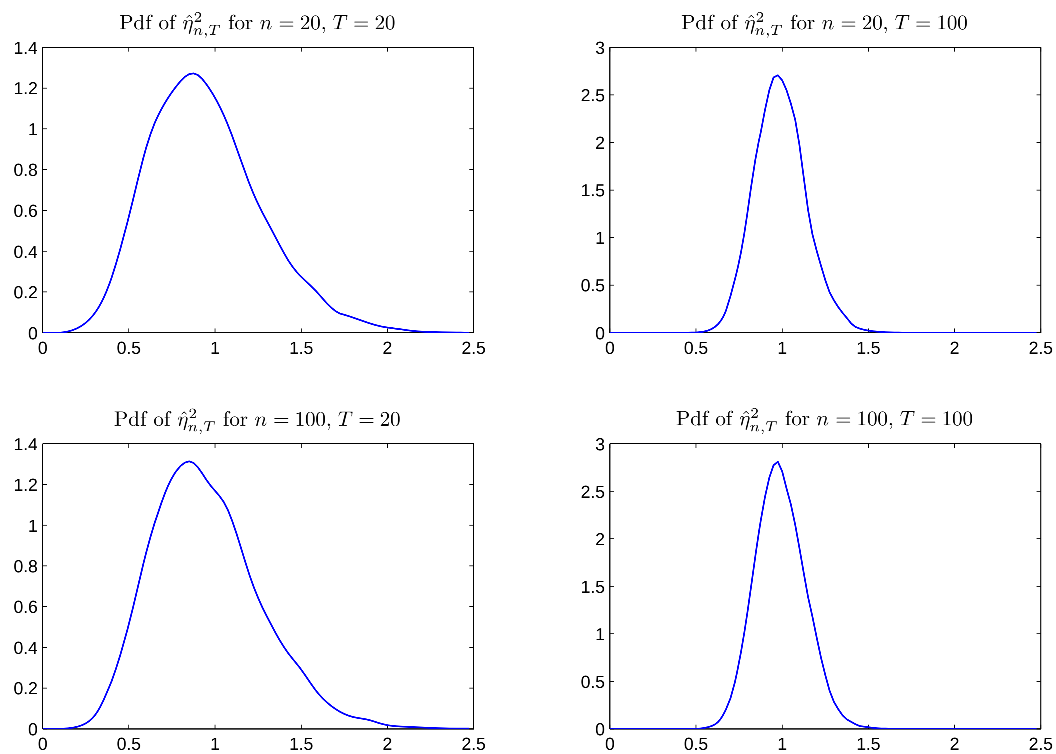
The Figure displays the pdf of estimator $\hat{\mu}_{n,T}$ for different sample sizes, that are $n = T = 20$ in the upper left panel, $n = 20$, $T = 100$ in the upper right panel, $n = 100$, $T = 20$ in the lower left panel and $n = T = 100$ in the lower right panel. The true values of the parameters are $\mu = 0$ and $\sigma^2 = \eta^2 = 1$.

Figure 2.2: Pdf of estimator $\hat{\sigma}^2_{n,t}$.



The Figure displays the pdf of estimator $\hat{\sigma}^2_{n,T}$ for different sample sizes, that are $n = T = 20$ in the upper left panel, $n = 20$, $T = 100$ in the upper right panel, $n = 100$, $T = 20$ in the lower left panel and $n = T = 100$ in the lower right panel. The true values of the parameters are $\mu = 0$ and $\sigma^2 = \eta^2 = 1$.

Figure 2.3: Pdf of estimator $\hat{\eta}^2_{n,t}$.



The Figure displays the pdf of estimator $\hat{\eta}^2_{n,T}$ for different sample sizes, that are $n = T = 20$ in the upper left panel, $n = 20$, $T = 100$ in the upper right panel, $n = 100$, $T = 20$ in the lower left panel and $n = T = 100$ in the lower right panel. The true values of the parameters are $\mu = 0$ and $\sigma^2 = \eta^2 = 1$.

Chapter 3

Static Qualitative Factor Model

This chapter proposes a unified setting for static factor models applied to panels of qualitative observations. We first describe in Section 3.1 the Single Risk Factor (SRF) model suggested in Basel 2 regulation for the analysis of default correlation [BCBS (2001)]. This model is a probit model with a common Gaussian factor. In Section 3.2, we consider a general qualitative model with Gaussian factors and macro-parameters only. Then, we explain how to get the CSA maximum likelihood estimator and GA estimator with adjustment for the variance, and derive their asymptotic properties. In some special cases the estimators and their asymptotic variances have closed form expressions. These models are discussed in Section 3.3. Finally, the results are applied to more complicated settings, such as stochastic intensity factor model (in Section 3.4), or factor analysis of dependence between qualitative variables (Section 3.5). Proofs are gathered in Appendix 3.6.

3.1 The Single Risk Factor Model for Default

This model has been initially introduced by Vasicek (1991) and is based on Merton's structural model [Merton (1974)].

i) The structural model

The structural model defines the default of a corporation from a (crude) analysis of its balance sheet. Let us denote by i, for $i = 1, \ldots, n$, the corporation assumed to be alive at the beginning of period $(t, t+1)$. The amount of debt to be reimbursed at the end of the period is known at date t and denoted $L_{i,t}$ (L for **liability**). The future **asset value** $A_{i,t+1}$ is uncertain. Then, the corporation defaults at $t+1$ if, and only if, the amount of asset is not sufficient to pay the debt, that is, if $A_{i,t+1} < L_{i,t}$. Thus, the default indicator is:

$$\begin{aligned} Y_{i,t+1} &= 1, \text{ if } A_{i,t+1} < L_{i,t}, \\ &= 0, \text{ otherwise,} \end{aligned}$$

or equivalently:

$$Y_{i,t+1} = \mathbb{1}_{\log A_{i,t+1} < \log L_{i,t}}, \tag{3.1}$$

where $\mathbb{1}$ denotes the indicator function.

If the log-asset value is Gaussian with mean $m_{A,i,t}$ and variance $\sigma^2_{A,i,t}$, conditional on the information available at time t, the conditional distribution of the default indicator is a Bernoulli distribution with parameter:

$$P_t[Y_{i,t+1} = 1] = \Phi\left[\frac{\log L_{i,t} - m_{A,i,t}}{\sigma_{A,i,t}}\right]. \tag{3.2}$$

The probability of default depends on the debt amount, on the expected log-asset value and on its volatility.

ii) The Single Risk Factor (SRF) Model

Merton's structural model is the basis for the specification proposed by Vasicek (1991), which concerns jointly n firms and allows for default correlation. In the original model, it is assumed that the n firms are identical. We describe below an extension in which the set of companies can be partitioned into K homogenous subpopulations, or cohorts, indexed by $k = 1, \ldots, K$.

We characterize by a double index (i, k) the corporation i in cohort k, for $i = 1, \ldots, n_k$ and $k = 1, \ldots, K$. In cohort k, the latent model for the log asset/liability ratio is:

$$\log A_{i,k,t+1} - \log L_{i,k,t} = a_k + b_k F_t + u_{i,k,t}, \tag{3.3}$$

where the variables F_t and $u_{i,k,t}$, with i, k, t varying, are independent, such that $F_t \sim N(0,1)$ and $u_{i,k,t} \sim N(0, \sigma_k^2)$. The variable F_t is a common factor representing systematic risk, while the errors $u_{i,k,t}$ correspond to idiosyncratic (or unsystematic) risks. The parameters a_k, b_k and σ_k are equal for all firms within cohort k, but may differ across cohorts. From (3.3) we deduce that the individual default indicators are independent conditional on the factor path, with conditional default probability:

$$PD_{k,t} = P[Y_{i,k,t+1} = 1 | F_t] = \Phi\left(-\frac{a_k + b_k F_t}{\sigma_k}\right). \tag{3.4}$$

The conditional default probability is stochastic and driven by the systematic factor F_t.

As usual in a dichotomous qualitative model, the parameters are identifiable up to a positive scaling factor. Equivalently, identifiable functions of the structural parameters [1] are $\alpha_k = -a_k/\sigma_k$ and $\beta_k = -b_k/\sigma_k$, say. Then the model becomes:

$$P[Y_{i,k,t+1} = 1 | F_t] = \Phi(\alpha_k + \beta_k F_t). \tag{3.5}$$

Remark 3.1: An alternative parameterization is proposed in the documents of the Basel Committee [see BCBS (2001), (2003)]. Since the unconditional distribution of the log asset-to-liability ratio $\log(A_{i,k,t+1}/L_{i,k,t})$ is Gaussian with mean a_k and variance $\sigma_k^2 + b_k^2$, the unconditional probability of default (PD) in cohort k is: [2]

$$PD_k = P(Y_{i,k,t} = 1) = \Phi\left(-\frac{a_k}{\sqrt{\sigma_k^2 + b_k^2}}\right).$$

[1] Up to the sign for β_k.

[2] The unconditional default probability PD_k is different from the conditional default probability equal to $PD_{k,t} = \Phi(\alpha_k + \beta_k F_t)$ (see also Figure 3.1).

Moreover, the correlation between the log(asset/liability) of two firms in a same cohort, called **asset correlation**, is: [3]

$$Corr[\log(A_{i,k,t+1}/L_{i,k,t}), \log(A_{j,k,t+1}/L_{j,k,t})] = b_k^2/(b_k^2 + \sigma_k^2) = \rho_k > 0, \text{ say.}$$

Thus, we get:

$$\begin{aligned} P(Y_{i,k,t+1} = 1|F_t) &= \Phi\left[-\frac{a_k}{\sigma_k} - \frac{b_k}{\sigma_k}F_t\right] \\ &= \Phi\left[\frac{\Phi^{-1}(PD_k)}{\sigma_k/\sqrt{\sigma_k^2 + b_k^2}} - \frac{b_k}{\sigma_k}F_t\right] \\ &= \Phi\left(\frac{\Phi^{-1}(PD_k) - \sqrt{\rho_k}F_t}{\sqrt{1-\rho_k}}\right). \end{aligned} \tag{3.6}$$

This new parameterization through the unconditional probability of default PD_k and asset correlation ρ_k is interesting for financial interpretation, although it is less convenient than the initial parameterization (3.5) for estimation purpose. Formula (3.6) shows how the conditional probability of default stochastically varies in time around its historical mean equal to PD_k. This is illustrated in Figure 3.1 for a simulated path of the factor F_t in a cohort of firms with unconditional default probability $PD = 0.05$. We consider two different values of asset correlation, that are $\rho = 0.10$ and $\rho = 0.30$, respectively.

[Insert Figure 3.1: Time-varying conditional default probability]

The conditional default probability features peaks at dates with large negative shocks in the factor. The time variability of the conditional default probability is more pronounced for larger values of the asset correlation.

Specification (3.5) can be written in a hierarchical way as:

$$P_t[Y_{i,k,t+1} = 1] = \Phi(a_{k,t}), \text{with } a_{k,t} = \alpha_k + \beta_k F_t, \quad F_t \sim IIN(0,1),$$

[3] In the Basel documents, the asset correlation is denoted ρ_k^2 instead of ρ_k.

or equivalently as:

$$\left\{ \begin{array}{l} P_t(Y_{i,k,t+1} = 1) = \Phi(a_{k,t}), \\ \\ \text{where vectors } a_t = (a_{1,t}, \ldots, a_{K,t})', \ t \text{ varying, are independent,} \\ \\ \text{with distribution } N(\alpha, \beta\beta'), \ \alpha = (\alpha_1, \ldots, \alpha_K)', \beta = (\beta_1, \ldots, \beta_K)'. \end{array} \right. \tag{3.7}$$

System (3.7) defines the **canonical factors** a_t and introduces restrictions on their distribution. These restrictions correspond to an exact factor structure for the components of vector a_t induced by the **reduced factor** F_t.

In the next two subsections we introduce simple estimation methodologies for parameters α and β.

iii) CSA estimator

Due to the homogeneity within cohorts, the individual observations can be summarized by the default frequencies:

$$\bar{Y}_{k,t+1} = \frac{1}{n_k} \sum_{i=1}^{n_k} Y_{i,k,t+1}. \tag{3.8}$$

To get the intuition for the CSA estimator, let us consider for a moment the (virtual) limiting case where the cohorts have infinite size, that is, $n_k = \infty$, $\forall k = 1, \ldots, K$. Then, the cross-sectional default frequencies are equal to the conditional probabilities of default:

$$\bar{Y}_{k,t+1} = E_t(Y_{i,k,t+1}) = PD_{k,t} = \Phi(a_{k,t}),$$

and the values of the canonical factors are known equal to:

$$a_{k,t} = \hat{a}_{k,t} = \Phi^{-1}(\bar{Y}_{k,t+1}). \tag{3.9}$$

We have to distinguish between the case with a single cohort, namely $K = 1$, which corresponds to the original Vasicek (1991) model, and the extension with $K \geq 2$ cohorts. In the first case, from (3.7) the canonical factor is such that:

$$a_t \sim IIN(\alpha, \beta^2). \tag{3.10}$$

Then, (3.9) and (3.10) suggest that the scalar parameters α and β can be estimated by ML applied to the time series of estimated canonical factors $\hat{a}_t$, to get:

$$\begin{aligned}
\hat{\alpha} &= \bar{a} = \frac{1}{T}\sum_{t=1}^{T} \hat{a}_t = \frac{1}{T}\sum_{t=1}^{T} \Phi^{-1}(\bar{Y}_{k,t+1}), \\
\hat{\beta}^2 &= \frac{1}{T}\sum_{t=1}^{T} (\hat{a}_t - \bar{a})^2 = \frac{1}{T}\sum_{t=1}^{T} \left(\Phi^{-1}(\bar{Y}_{k,t+1}) - \frac{1}{T}\sum_{t=1}^{T} \Phi^{-1}(\bar{Y}_{k,t+1}) \right)^2 .
\end{aligned}$$

These estimators are called CSA estimators. Although the CSA estimators have been motivated by the limiting argument of infinite cohort sizes, they can be computed with finite cohort sizes and are expected to yield rather accurate estimates when the cohort sizes are sufficiently large. The large sample properties of the CSA estimators are discussed in Section 3.2 iii) for the general model.

When we have more than one cohort ($K \geq 2$), a similar approach cannot be followed since the distribution of the canonical factors in (3.7) is degenerate with a singular variance-covariance matrix. This is because we have a linear deterministic relationship between the canonical factors. Indeed, in the limiting case of infinite cohort sizes, we could deduce without error the values of parameter vectors α, β and factor values F_t, $t = 1, \ldots, T$ by solving the system of KT equations:

$$a_{k,t} = \alpha_k + \beta_k F_t, \quad k = 1, \ldots, \ K, \ t = 1, \ldots, T,$$

in the $2K+T$ unknown quantities. When the estimation error for the canonical factors is taken into account, it is seen in the next subsection that a non-degenerate log-likelihood function is recovered. However, with more than one cohort, it is natural to include cohort-specific effects in the canonical factors. Then, the distribution of the canonical factors becomes non-singular and a well defined CSA estimator can be derived [see Section 3.3 ii) and in particular the discussion in Remark 3.3 for the financial relevance of including cohort specific effects].

iv) Variance Granularity Adjusted (VGA) estimators

Since in reality the cohort sizes are large, but finite, we may expect that the CSA approach can be improved by taking into account the estimation error on the canonical factors. By applying the CLT by date and cohort, we see that the default frequencies $\bar{Y}_{k,t+1}$ are asymptotically independent, with mean $PD_{k,t} = \Phi(a_{k,t})$ and variance $\dfrac{PD_{kt}(1-PD_{kt})}{n_k} = \dfrac{\Phi(a_{k,t})[1-\Phi(a_{k,t})]}{n_k}$.

By applying the delta method and noting that the derivative of function $\Phi^{-1}(.)$ is $1/\varphi[\Phi^{-1}(.)]$, where φ denotes the pdf of the standard normal distribution, we deduce that the approximations of the canonical factors are also independent and asymptotically Gaussian:

$$\hat{a}_{k,t} = \Phi^{-1}(\bar{Y}_{k,t+1}) \simeq N\left(a_{k,t}, \frac{\Phi(a_{k,t})[1-\Phi(a_{k,t})]}{n_k \varphi(a_{k,t})^2}\right).$$

Equivalently:

$$\begin{aligned}\hat{a}_{k,t} &\simeq a_{k,t} + \left(\frac{\Phi(a_{k,t})[1-\Phi(a_{k,t})]}{n_k \varphi(a_{k,t})^2}\right)^{1/2} v_{k,t} \\ &\simeq \alpha_k + \beta_k F_t + \left(\frac{\Phi(\hat{a}_{k,t})[1-\Phi(\hat{a}_{k,t})]}{n_k \varphi(\hat{a}_{k,t})^2}\right)^{1/2} v_{k,t},\end{aligned}$$

where F_t and $v_{k,t}$, k, t varying, are independent, standard Gaussian variables. Let us denote by Δ_t the $K \times K$ diagonal matrix with elements $\Phi(\hat{a}_{k,t})[1-\Phi(\hat{a}_{k,t})]/[n_k \varphi(\hat{a}_{k,t})^2]$, $k = 1, \ldots, K$. The parameters α and β will be estimated by optimizing the VGA log-likelihood function:

$$\begin{aligned}L^{VGA}(\alpha, \beta) &= \sum_{t=1}^{T} \left\{ -\frac{K}{2}\log(2\pi) - \frac{1}{2}\log\det(\beta\beta' + \Delta_t) \right. \\ &\quad \left. - \frac{1}{2}(\hat{a}_t - \alpha)'(\beta\beta' + \Delta_t)^{-1}(\hat{a}_t - \alpha) \right\}. \qquad (3.11)\end{aligned}$$

These estimators are called **Variance Granularity Adjusted (VGA) maximum likelihood** estimators. In order to take into account the finite cross-sectional size, we have introduced an adjustment of the variance of the error term, which explains the terminology.

Remark 3.2: The VGA maximum likelihood method has to be compared with the finite sample ML method. The true log-likelihood function is:

$$\begin{aligned} & L(\alpha, \beta) \\ = \; & \sum_{t=1}^{T} \log \left[\int \prod_{k=1}^{K} \left\{ \Phi(\alpha_k + \beta_k f)^{n_{k,t}} [1 - \Phi(\alpha_k + \beta_k f)]^{n_k - n_{k,t}} \right\} \frac{1}{\sqrt{2\pi}} \exp(-f^2/2) df \right], \end{aligned} \tag{3.12}$$

where $n_{k,t} = \sum_{i=1}^{n_k} Y_{i,k,t} = n_k \bar{Y}_{k,t}$ is the number of defaults in cohort k for period $(t-1, t)$. When the true log-likelihood is maximized, the integrals in (3.12) are often approximated by simulation, leading to simulated maximum likelihood estimators [see e.g. Gouriéroux, Monfort (1996)]. The approximation (3.11) circumvents the computation of the T integrals involved in (3.12). We will see in Chapter 4 that function $L^{VGA}(\alpha, \beta)$ can be derived from an asymptotic expansion of $L(\alpha, \beta)$ when the cohort sizes n_k are large, and the estimators obtained by maximizing $L^{VGA}(\alpha, \beta)$ are asymptotically equivalent to the ML estimator. Moreover, we will see that it can be appropriate to introduce a granularity adjustment for the mean too.

3.2 The General Model and its Estimation

The approaches described for the SRF model can be extended to more general static qualitative factor models.

i) The model

As in the SRF model for default, let us consider a set of cohorts and individual observations of a qualitative variable $Y_{i,k,t}$, for $i = 1, \ldots, n_k$,

$k = 1, \ldots, K$, $t = 1, \ldots, T$. The qualitative variable is polytomous with J alternatives [4].

The model is defined in two steps. We first explain how the distribution of the observations depends on underlying canonical factors; then, restrictions on the canonical factor distribution are introduced.

(*) Distribution of the observations given the canonical factors

The individual observations are assumed independent, conditionally on canonical factors $a_{k,t}$, $k = 1, \ldots, K$, $t = 1, \ldots, T$:

$$P[Y_{i,k,t} = j | a_t] = p(j; a_{k,t}), \tag{3.13}$$

where $p(j; .)$ denotes the elementary probability of alternative j, for $j = 1, \ldots, J$. The distribution can depend on cohort and time by means of the canonical factor, but does not depend on the individual within the cohort. The canonical factor can be multidimensional, with dimension $\dim(a_{k,t}) = S$, say, and we assume that it can take any value in $I\!R^S$.

(**) Joint distribution of the canonical factors

The model is completed by specifying the distribution of the canonical factors. Let us introduce the KS-dimensional vector of canonical factor values at date t, denoted by $a_t = (a'_{1,t}, \ldots, a'_{K,t})'$. We assume that the random vectors a_t, $t = 1, \ldots, T$, are independent with identical Gaussian distributions:

$$a_t \sim IIN[\mu(\theta), \Omega(\theta)], \tag{3.14}$$

where θ is a p-dimensional unknown parameter, and matrix $\Omega(\theta)$ is invertible. The model is static due to the assumption of serial independence of the factors.

[4] Recall that a one-dimensional polytomous variable with J alternatives can be equivalently represented as a J-dimensional vector of dichotomous qualitative components. The components are the indicators of the J different alternatives and sum up to one.

The likelihood function of model (3.13)-(3.14) is:

$$\begin{aligned} l(\underline{y_T};\theta) &= \prod_{t=1}^{T} \int \cdots \int \prod_{k=1}^{K} \prod_{i=1}^{n_k} p(y_{i,k,t}; a_{k,t}) \\ & \frac{1}{(2\pi)^{SK/2}[\det \Omega(\theta)]^{1/2}} \exp\left\{ -\frac{1}{2}[a_t - \mu(\theta)]'\Omega(\theta)^{-1}[a_t - \mu(\theta)] \right\} da_t, \end{aligned} \tag{3.15}$$

where $\underline{y_T}$ denotes the individual histories $y_{i,1}, \ldots, y_{i,T}$ for $i = 1, \ldots, n$. The likelihood function depends on macro-parameter θ and involves multidimensional integrals with dimension KS. Since the cohorts are homogenous, the likelihood function can be simplified. Let us denote by $n_{j,k,t}$ the number of observations taking alternative j, in cohort k, at time t. We get:

$$\begin{aligned} l(\underline{y_T};\theta) &= \prod_{t=1}^{T} \int \cdots \int \prod_{k=1}^{K} \prod_{j=1}^{J} p(j; a_{k,t})^{n_{j,k,t}} \\ & \frac{1}{(2\pi)^{SK/2}[\det \Omega(\theta)]^{1/2}} \exp\left\{ -\frac{1}{2}[a_t - \mu(\theta)]'\Omega(\theta)^{-1}[a_t - \mu(\theta)] \right\} da_t. \end{aligned} \tag{3.16}$$

Thus, without loss of information, the cross-sectional observations for cohort k can be summarized by the J cross-sectional aggregates $n_{j,k,t}, j = 1, \ldots, J$.

The likelihood function (3.16) is complicated because of the T numerical integrals of dimension KS. We consider below estimators of θ that are computationally simpler than the Maximum Likelihood (ML) estimator.

ii) The fixed effect maximum likelihood estimator

By analogy with the discussion of the SFR model in Section 3.1, let us consider the cross-sectional observations for a given date t, and treat a_t as an unknown parameter. Approximate factor values are the **fixed effects ML**

estimators defined by:

$$\hat{a}_{k,t} = \arg\max_{a_{k,t}} \sum_{i=1}^{n_k} \log p(y_{i,k,t}; a_{k,t}) \tag{3.17}$$

$$= \arg\max_{a_{k,t}} \sum_{j=1}^{J} n_{j,k,t} \log p(j; a_{k,t}), \tag{3.18}$$

where the arg max operator provides the argument $a_{k,t}$ that maximizes the objective function.

Identification assumptions have to be introduced to ensure a unique solution to the cross-sectional optimization above. Intuitively, we must have less "parameters" than (linearly independent) aggregate observations, that is the order condition:

$$S \leq J - 1. \tag{3.19}$$

When $S = J - 1$, the canonical factors are just-identified; they are overidentified, if $S < J - 1$.

The LLN and CLT can be applied conditionally on the canonical factor values, if n_k is large for any $k = 1, \ldots, K$. Hence, the standard asymptotic results for maximum likelihood estimators are valid. More precisely, the fixed effect ML estimators $\hat{a}_{k,t}$, for $k = 1, \ldots, K$, $t = 1, \ldots, T$ are asymptotically independent, with:

$$\sqrt{n_k}(\hat{a}_{k,t} - a_{k,t}) \xrightarrow{d} N(0, \Sigma_{k,t}), \tag{3.20}$$

where

$$\Sigma_{k,t} = \left\{ E \left[-\frac{\partial^2 \log p(Y_{i,k,t}; a_{k,t})}{\partial a \partial a'} | a_{k,t} \right] \right\}^{-1}. \tag{3.21}$$

The asymptotic variance of $\hat{a}_{k,t}$ is the inverse of an information matrix computed as if $a_{k,t}$ were a (multidimensional) parameter. The derivatives are taken with respect to $a_{k,t}$ and the computation of the expectation is performed conditional on $a_{k,t}$, that is, as if $a_{k,t}$ were a vector of constants.

iii) The CSA maximum likelihood estimator

The motivation for the CSA estimator is best understood if we consider for a moment the limiting (virtual) case where the cohort sizes are infinite, that is, $n_k = \infty$ for $k = 1, \ldots, K$. Then, the fixed effects ML estimators would coincide with the unknown canonical factor values by the LLN. The log-likelihood function would become:

$$L^{\text{CSA}}(\theta) \propto -\frac{T}{2} \log \det \Omega(\theta) - \frac{1}{2} \sum_{t=1}^{T} [\hat{a}_t - \mu(\theta)]' \Omega(\theta)^{-1} [\hat{a}_t - \mu(\theta)]. \tag{3.22}$$

This argument suggests to consider the CSA maximum likelihood estimator of θ defined by:

$$\hat{\theta}^{\text{CSA}} = \arg\max_{\theta} L^{\text{CSA}}(\theta). \tag{3.23}$$

Let us now discuss the asymptotic distribution of the CSA estimator when both the cross-sectional dimension T and the time dimension n are large ($n, T \to \infty$). When the cross-sectional dimension n is much larger than T (i.e. $T/n \to 0$), the large sample distribution of $\hat{\theta}^{\text{CSA}}$ is the same as if the canonical factors were observable $a_t = \hat{a}_t$, and we can apply the standard asymptotic theory with respect to time ($T \to \infty$) for the log-likelihood function (3.22) [see references in Chapter 4 for the regularity conditions]. We deduce that the CSA estimator is consistent, at speed $1/\sqrt{T}$, with asymptotic distribution:

$$\sqrt{T}(\hat{\theta}^{\text{CSA}} - \theta) \xrightarrow{d} N\left(0, \left[\underset{T\to\infty}{plim} - \frac{1}{T} \frac{\partial^2 L^{\text{CSA}}(\theta)}{\partial\theta \partial\theta'}\right]^{-1}\right). \tag{3.24}$$

In particular, the CSA estimator $\hat{\theta}^{\text{CSA}}$ is asymptotically equivalent to the true ML estimator of θ that maximizes the likelihood (3.16). For the SRF model with $K = 1$ cohort [see Section 3.1 iii)], the asymptotic variances of the CSA estimators are $AsVar(\hat{\alpha}) = \frac{1}{T}\beta^2$ and $AsVar(\hat{\beta}^2) = \frac{2}{T}\beta^4$, and the estimators of α and β are asymptotically independent.

iv) The VGA maximum likelihood estimator

The VGA estimator accounts for the difference between the fixed effects estimates and the true factor values when the cohort sizes are large, but finite. From (3.20) we deduce that:

$$\hat{a}_{k,t} \simeq a_{k,t} + \frac{1}{\sqrt{n_k}} \Sigma_{k,t}^{1/2} v_{k,t}, \tag{3.25}$$

where the errors (v_{kt}) are standard normal, independent of each other, and independent of the $a_{k,t}$'s. Therefore, by integrating out the unobservable canonical factors, we get:

$$\hat{a}_t \approx N[\mu(\theta), \Omega(\theta) + \hat{\Sigma}_{n,t}], \tag{3.26}$$

where $\hat{\Sigma}_{n,t} = \text{diag}\,[\hat{\Sigma}_{kt}/n_k]$, and $\hat{\Sigma}_{kt}$ is a consistent estimator of Σ_{kt}.

The variance granularity adjusted log-likelihood function is:

$$L^{\text{VGA}}(\theta) \propto -\frac{1}{2} \sum_{t=1}^{T} \log\det[\Omega(\theta) + \hat{\Sigma}_{n,t}] \tag{3.27}$$

$$-\frac{1}{2} \sum_{t=1}^{T} [\hat{a}_t - \mu(\theta)]' [\Omega(\theta) + \hat{\Sigma}_{n,t}]^{-1} [\hat{a}_t - \mu(\theta)]. \tag{3.28}$$

Compared to the CSA log-likelihood function (3.22), the variance has been adjusted to account for the variability of the fixed effects estimators of the canonical factors. The VGA maximum likelihood estimator of θ is:

$$\hat{\theta}^{\text{VGA}} = \arg\max_{\theta} L^{\text{VGA}}(\theta). \tag{3.29}$$

For large n and T $(n, T \to \infty,\ T/n \to 0)$, the asymptotic distribution of the VGA estimator is the same as the one of the CSA estimator given in (3.24). In particular, the VGA estimator is consistent at speed $1/\sqrt{T}$ and asymptotically normal. The VGA estimator differs from the CSA estimator in terms of higher order asymptotic properties, more specifically, in terms of the bias at order $1/n$ (see Chapter 4, Section 4.3).

3.3 Closed Form Expressions of the Estimators

The CSA and VGA log-likelihood functions have closed form expressions, and, in particular, they do not involve the multiple integrals appearing in the finite sample likelihood function [see equation (3.16)]. In important special cases of static qualitative model with factors, it is possible to get also closed form expressions of the CSA maximum likelihood estimators themselves, and of their asymptotic variance [see Gouriéroux, Monfort (2010)]. We describe below such simplifications.

i) Just-identified canonical factors

Let us denote by $p_{j,k,t} = p(j; a_{k,t})$ the true elementary probabilities and by $p_{k,t} = (p_{1,k,t}, \dots, p_{J,k,t})'$ the associated vector of probabilities for cohort k and time t. Under the assumption of just identification $S = J - 1$, we can write:

$$p_{k,t} = \Pi(a_{k,t}), \tag{3.30}$$

where Π is a one-to-one function of $I\!R^{J-1}$ onto the simplex of $I\!R^J$, that is, the set of discrete probability distributions:

$$\left\{ (p_1, \dots, p_J)', \text{ with } p_j \geq 0, j = 1, \dots, J, \sum_{j=1}^{J} p_j = 1 \right\}.$$

In several examples (see below and Sections 3.4, 3.5), function Π can be inverted to express the canonical factors in terms of elementary probabilities as:

$$a_{k,t} = c(p_{k,t}), \quad \text{say.} \tag{3.31}$$

It is easily checked that the solution of optimization (3.18) is such that:

$$\hat{p}_{k,t} = \Pi(\hat{a}_{k,t}), \tag{3.32}$$

where $\hat{p}_{k,t} = (n_{1,k,t}/n_k, \dots, n_{J,k,t}/n_k)'$ are the observed cross-sectional frequencies of the alternatives at date t. We deduce the closed form expression of the fixed effects ML estimators of the canonical factors:

$$\hat{a}_{k,t} = c(\hat{p}_{k,t}). \tag{3.33}$$

It follows that [see e.g. Gouriéroux, Monfort (1989), Example 7.19, and the δ-method]:

$$\sqrt{n}_k(\hat{a}_{k,t} - a_{k,t}) \xrightarrow{d} N\left(0, \frac{\partial c(p_{k,t})}{\partial p'_{k,t}}[\text{diag}(p_{k,t}) - p_{k,t}p'_{k,t}]\frac{\partial c(p_{k,t})'}{\partial p_{k,t}}\right), \tag{3.34}$$

and that the variance-covariance matrix $\Sigma_{k,t}$ in (3.21) is consistently estimated by:

$$\hat{\Sigma}_{k,t} = \frac{\partial c(\hat{p}_{k,t})}{\partial p'_{k,t}}(\text{diag}\ \hat{p}_{k,t} - \hat{p}_{k,t}\hat{p}'_{k,t})\frac{\partial c(\hat{p}_{k,t})'}{\partial p_{k,t}}. \tag{3.35}$$

To summarize the result above, the derivations of the fixed effects ML estimator of the canonical factor and of their estimated asymptotic variance become simple, if the canonical factor can be interpreted as a reparameterization of the qualitative model by $J-1$ real parameters. Such real parameterization have often been considered in the literature on qualitative models for the purpose of introducing quantitative exogenous variables. We review below such standard reparameterizations for the main qualitative models. For expository purpose, we keep the index j, but omit indexes k and t. We provide for each example the function Π and the function c.

Example 3.1: Dichotomous probit model ($J = 2$)

$$p_1 = \Phi(a_1), \quad p_2 = 1 - \Phi(a_1).$$

This corresponds to the Merton (1987) - Vasicek (1991) default model described in Section 3.1. The probability p_1 is displayed as a function of the canonical factor a_1 in Figure 3.2.

[Insert Figure 3.2: Parameterization of the dichotomous probit model]

This mapping from $I\!R$ to $(0,1)$ is one-to-one. We deduce $a_1 = \Phi^{-1}(p_1)$, where Φ^{-1} is the quantile function of the standard normal distribution, called probit function.

Example 3.2: Dichotomous logit model ($J = 2$)

$$p_1 = [1 - \exp(-a_1)]^{-1}, \quad p_2 = \exp(-a_1)[1 + \exp(-a_1)]^{-1}.$$

The probability p_1 is displayed as a function of the canonical factor a_1 in Figure 3.3.

[Insert Figure 3.3: Parameterization of the dichotomous logit model]

We deduce $a_1 = \log[p_1/(1 - p_1)]$ and function c is the inverse of the logistic function, that is the logit function [see Berkson (1944)].

Example 3.3: Multinomial logit model [McFadden (1973), (1976), any $J \geq 2$].

The model is reparameterized as:

$$p_j = \exp(a_j)[\sum_{l=1}^{J} \exp(a_l)]^{-1}, \quad j = 1, \dots, J,$$

with the convention $a_1 = 0$. We get:

$$a_j = \log(p_j/p_1), \quad j = 2, \dots, J.$$

The canonical factors are log transforms of appropriate **odd ratios**. For the case of $J = 3$ alternatives, we display in Figure 3.4 the canonical factors a_2 and a_3 as functions of the elementary probabilities (p_1, p_2, p_3).

[Insert Figure 3.4: Canonical factors in the multinomial logit model]

The vector function (a_2, a_3) is a one-to-one mapping from the simplex in $I\!R^3$ onto $I\!R^2$. The component function a_2 admits large positive (resp. negative) values close to the boundary of the simplex with $p_2 = 0$ and $p_1 > 0$ (resp. $p_1 = 0$ and $p_2 > 0$). We have $a_2 = 0$ on the intersection of the simplex with the plane $p_1 = p_2$, which corresponds to the lightest part of the coloured surface. The point $p_1 = 0$, $p_2 = 0$, $p_3 = 1$ is singular, since function a_2 can admit any real value in a neighbourhood of this point. The component function a_3 features a similar behaviour interchanging p_3 with p_2. For instance, the lightest part of the coloured surface is obtained for $p_1 = p_3$, that is, $2p_1 + p_2 = 1$.

Example 3.4: Ordered polytomous probit model (any $J \geq 2$).

The model is reparameterized as:

$$p_j = \Phi(a_j) - \Phi(a_{j-1}), \quad j = 1, \ldots, J,$$

with the convention $a_0 = -\infty, a_J = +\infty$. We deduce:

$$a_j = \Phi^{-1}(p_1 + \ldots + p_j). \tag{3.36}$$

This new parameterization does not completely fulfill our assumptions. Indeed, parameters a_j, $j = 1, \ldots, J$, are real, but constrained to form an increasing sequence. However, the Gaussian assumption on the canonical factor [see (3.14)] is still relevant, if it concerns parallel shift on the canonical factors only, that is, if we write:

$$a_{jt} = \alpha_j + \beta F_t, \tag{3.37}$$

with $\alpha_1 \leq \alpha_2 \leq \ldots \leq \alpha_{J-1}$, where F_t is a Gaussian random variable. Indeed, since scalar parameter β is independent of the alternative, the ordering of the intercepts implies the similar ordering for the canonical factors.

The Gaussian distribution of the random vector a_t implied by (3.37) is degenerate, because of the deterministic relationships between the canonical factors associated with the different alternatives. We have already encountered a similar feature in the SRF model with several cohorts in Section 3.1. Due to this degeneracy, the results of Section 3.2 on the rate of convergence of the estimators do not apply. In particular, some of the parameters among α_j and β have a micro-interpretation, and feature a convergence rate $1/\sqrt{nT}$. The estimators of parameters α_j and β and their asymptotic properties can be derived by using the results presented in Chapter 4, where we consider models with both macro- and micro-parameters (see in particular Section 4.3 on rating migration models based on ordered qualitative specifications).

ii) Gaussian factor analysis of the canonical factors

In order to structure the cross dependence between the canonical factors, let us introduce a linear Gaussian factor model for the distribution of a_t. The model is defined by:

$$a_t = \alpha + \beta F_t + \eta w_t, \quad t = 1, \dots, T, \tag{3.38}$$

where α [resp. β] is a vector of dimension KS [resp. a matrix of dimension (KS, L)], η is a positive scalar, (F_t) are independent Gaussian vectors $F_t \sim IIN(0, Id)$ with size $L < KS$, and w_t are independent standard Gaussian vectors with size KS, that is, $w_t \sim IIN(0, Id)$. Moreover, the factors (F_t) and the errors (w_t) are independent. We deduce that the common distribution of the canonical factor is:

$$a_t \sim N(\alpha, \beta\beta' + \eta^2 Id). \tag{3.39}$$

Thus, model (3.38) implies a special structure on the variance-covariance matrix of the canonical factors. Indeed, the cross-covariances are captured by means of a matrix $\beta\beta'$ of reduced rank L, where L is the number of underlying static factors.

Remark 3.3: It is important to compare the static factor model (3.38) with the latent factor model (3.3) usually introduced in the SRF model for default. Model (3.3) includes individual error terms $u_{i,k,t}$, whose effects vanish by cross-sectional aggregation. This explains, why the associated model for canonical factors reduces to $a_t = \alpha + \beta F_t$, that is, does not include the error terms w_t. When $K \geq 2$, a consequence is the non invertibility of the matrix $\Omega = V(a_t) = \beta\beta'$, and the degeneracy of the CSA likelihood function [see Section 3.1 iii)]. More importantly from a financial point of view, by implicitly setting $\eta = 0$, the basic SRF model neglects the cohort specific source of risk and consequently underestimates the required capital. Thus, with $K \geq 2$, it is preferable to include in the SRF model an additional error term w_t in the canonical factors as in (3.38), which corresponds to cohort-specific effects. Finally, when we have a single cohort ($K = 1$), it is not possible to introduce an additional error term w_t in the canonical factor, since the associated parameter η is not identified.

The vector of parameters in model (3.38) is $\theta = [\alpha', (vec\beta)', \eta]'$. However, it is known that the factors F and the factor sensitivities β are defined up to a linear orthogonal transformation. Therefore, without loss of generality, we can impose the identification restrictions:

Identification restrictions: $\beta_k'\beta_l = 0, \forall k \neq l$, *where* $\beta_l, l = 1, \ldots, L$, *denote the columns of matrix* β.

Let us now derive the CSA estimator of parameter θ. Let us denote:

$$\bar{a}_T = \frac{1}{T}\sum_{t=1}^{T} \hat{a}_t, \tag{3.40}$$

the historical mean of the estimated canonical factors, and:

$$\hat{V}_T = \frac{1}{T}\sum_{t=1}^{T} (\hat{a}_t - \bar{a}_T)(\hat{a}_t - \bar{a}_T)', \tag{3.41}$$

their historical variance-covariance matrix. The spectral decomposition of the historical variance-covariance matrix $\hat{V}_T$ provides a decreasing sequence of nonnegative eigenvalues $\hat{\lambda}_{1,T} \geq \hat{\lambda}_{2,T} \geq \cdots$, with associated orthonormal eigenvectors $\hat{e}_{1,T}$, $\hat{e}_{2,T}$, $\cdots$. The proposition below provides the explicit expressions of the CSA maximum likelihood estimator of parameter θ.

Proposition 3.3: *The CSA maximum likelihood estimators of the components of parameter θ are:*

$$\begin{aligned} \hat{\alpha}_T &= \bar{a}_T, \quad \hat{\eta}_T^2 = [Tr(\hat{V}_T) - \sum_{l=1}^{L} \hat{\lambda}_{l,T}]/(KS - L), \\ \hat{\beta}_{l,T} &= (\hat{\lambda}_{l,T} - \hat{\eta}_T^2)^{1/2} \hat{e}_{l,T}, \quad l = 1, \ldots, L. \end{aligned}$$

Proof: See Appendix 3.6.

These maximum likelihood estimators are based on the **Spectral Decomposition** of matrix $\hat{V}_T$ [see e.g. Anderson (2003) and Review A.4]. The asymptotic variance-covariance matrix has also an explicit expression [see e.g. Gouriéroux, Monfort (2010), Section 3.3][5].

iii) The estimation steps

Under the conditions of Subsections i) and ii) above, the estimation steps can be summarized as follows:

Step 1: Reparameterize the qualitative model in the appropriate way to get $a_t = c(p_t)$, with real values.

Step 2: Compute the observed frequencies $\hat{p}_t$ and deduce the estimated canonical factors as $\hat{a}_t = c(\hat{p}_t)$.

Step 3: Compute the historical mean $\bar{a}_T$ and variance $\hat{V}_T$ of the estimated canonical factors.

Step 4: Get the CSA estimates of α, β, η^2 from the spectral decomposition of matrix $\hat{V}_T$ (see Proposition 3.3).

Step 5: Finally, get the VGA estimates by optimizing numerically the VGA log-likelihood function with the CSA estimate as starting value of the optimization algorithm.

iv) Illustration: Factor model for corporate default

As an illustration of the above methodology, we estimate a factor model for corporate default. The binary variable $Y_{i,k,t}$ is a firm's default indicator ($J = 2$, see Section 3.1) and the cohorts $k = 1, 2, 3$ correspond to the non-investment-grade rating classes BB, B and C in the Standard & Poor's (S&P) rating system ($K = 3$). The series of 1-year default frequencies $\bar{Y}_{k,t}$ are displayed in Figure 3.5 for the period 1990-2009 ($T = 20$). These default frequencies are deduced from the S&P rating transition matrices, which are

[5]The initial derivation of these variances in Lawley, Maxwell (1971) provides only approximated variance-covariance matrices [see also Jennrich, Thayer (1977)].

computed from a large pool of US large and medium-size firms (see Section 4.5 for a more detailed description of the data). The cohort size n_k is of the order of thousand firms for rating classes BB and B, and of the order of some hundreds for rating class C.

[Insert Figure 3.5: S&P US corporate default frequencies for rating classes BB, B and C.]

As expected, at any given date the default frequencies are ranked in terms of the riskiness of the speculative rating class. Moreover, the series of default frequencies of the three rating classes feature a similar countercyclical pattern, with peaks of default intensity associated with recessions in the US economy (1990-91, 2001, and 2008-2009).

Since the risk variable is dichotomous, we have a single canonical factor for each rating class ($S = 1$), and the vector of canonical factors a_t is trivariate. We adopt a probit specification (see Example 3.1). The series of estimated canonical factor values $\hat{a}_{k,t} = \Phi^{-1}(\bar{Y}_{k,t})$ for the three rating classes are displayed in Figure 3.6.

[Insert Figure 3.6: Estimated canonical factor values for rating classes BB, B and C.]

The estimated canonical factor values are the quantiles of the standard Gaussian distribution for the percentiles that correspond to the default frequencies in Figure 3.5. Steps 1 and 2 of the estimation methodology are completed.

Let us now apply steps 3-4. The historical mean and variance of the estimated canonical factor vectors are:

$$\bar{a}_T = \begin{pmatrix} -2.423 \\ -1.671 \\ -0.566 \end{pmatrix}, \qquad \hat{V}_T = \begin{pmatrix} 0.126 & 0.094 & 0.085 \\ 0.094 & 0.174 & 0.129 \\ 0.085 & 0.129 & 0.200 \end{pmatrix}. \tag{3.42}$$

The SVD of matrix $\hat{V}_T$ is characterized by the 3 eigenvalues:

$$\hat{\lambda}_{1,T} = 0.379, \qquad \hat{\lambda}_{2,T} = 0.072, \qquad \hat{\lambda}_{3,T} = 0.049,$$

with associated orthonormal eigenvectors:

$$\hat{e}_{1,T} = \begin{pmatrix} 0.447 \\ 0.613 \\ 0.652 \end{pmatrix}, \qquad \hat{e}_{2,T} = \begin{pmatrix} 0.702 \\ 0.212 \\ -0.680 \end{pmatrix}, \qquad \hat{e}_{3,T} = \begin{pmatrix} 0.555 \\ -0.761 \\ 0.335 \end{pmatrix}.$$

The first eigenvalue $\hat{\lambda}_{1,T}$ of matrix $\hat{V}_T$ is significantly larger than the other two. Moreover, the components of the eigenvector associated with $\hat{\lambda}_{1,T}$ have the same sign across rating classes, while the eigenvectors associated with the other two eigenvalues have components of both signs. Intuitively, these findings are compatible with a single common factor having a similar impact on the default risk of the three rating classes. Hence, we use a Gaussian single-factor model for the canonical factors as in Equation (3.38) with $L = 1$.

Let us compute the CSA estimates of the model parameters. From Proposition 3.3, the CSA estimate of the vector of intercepts is $\hat{\alpha}_T = \bar{a}_T$ given in (3.42), while the CSA estimates of the vector of sensitivities β and idiosyncratic variance η^2 are:

$$\hat{\beta}_T = \begin{pmatrix} 0.252 \\ 0.346 \\ 0.368 \end{pmatrix}, \qquad \hat{\eta}_T^2 = 0.060.$$

As expected, the estimated intercepts are increasing w.r.t. the riskiness of the rating class. The sensitivities to the common factor have the same sign across rating classes, and are larger in magnitude for the riskiest rating classes B and C than for rating class BB. The sign of the eigenvector associated with the largest eigenvalue of $\hat{V}_T$ has been selected to get positive factor sensitivities and interpret the common factor as a default risk factor. From equation (3.39), the estimates of the unconditional variances of the canonical factors are 0.124, 0.180 and 0.195 for rating classes BB, B and C, respectively. Hence, for rating class BB the systematic factor and the idiosyncratic factor contribute each about an half of the unconditional variance of the canonical factor. For rating classes B and C the proportions are about 2/3 from the systematic factor and 1/3 from the idiosyncratic factor.

3.4 Stochastic Intensity Model with Factor

A discrete random variable with a fixed number K of admissible values can be identified with a polytomous qualitative variable by considering the set of values as the set of alternatives. This interpretation is especially interesting for duration variables representing the time to some given event, such as default or prepayment in credit analysis, and death or lapse for life insurance contracts.

i) Distribution of a duration variable

There exist alternative characterizations of the distribution of a duration variable Y, with values $k = 1, \ldots, K$. We can consider the elementary probabilities:

$$\pi_k = P[Y = k], \quad k = 1, \ldots, K.$$

We can also consider the successive **intensities** of event occurrence. These intensities measure the short term probability of occurrence of the event by means of the following conditional probabilities:

$$p_k = P[Y = k | Y \geq k], \quad k = 1, \ldots, K. \tag{3.43}$$

The elementary probabilities and the intensities are in a one-to-one relationship. Indeed, we have:

$$p_k = \pi_k / \sum_{l=k}^{K} \pi_l, \tag{3.44}$$

and:

$$\pi_k = \left[\prod_{l=1}^{k-1} (1 - p_l) \right] p_k. \tag{3.45}$$

There exist at least three advantages of an approach based on intensity. First, the intensities p_k, $k = 1, \ldots, K-1$, can be fixed independently between 0 and 1 (with $p_K = 1$), whereas the elementary probabilities are subject to the unit mass restriction. Second, the sample counterparts of the intensities in an i.i.d. framework are asymptotically independent [6], Gaussian with mean p_k and variance $p_k(1-p_k)/n_k$, where n_k is the number of individuals in the **Population-at-Risk** (PaR), i.e. with duration larger, or equal to k. Third, intensities allow for an appropriate treatment of two competing notion of times, that are calendar time with time origin Jesus-Christ birth and individual time with time origin the beginning of the contract.

ii) Stochastic intensity with factor

Let us now consider a large population of contracts originated at different consecutive dates. This population is assumed homogenous; in particular all the contracts have the same contractual term K, say. We assume that we monitor the contracts over a given period of time $t = 1, \ldots, T$, and observe whether they get closed before their contractual term, or not. Thus, at any given date t, we can observe K different categories of contracts still alive, depending if they have been originated at date $t-1, t-2, \ldots,$ or $t-K$. Among them, only the first $K-1$ categories can lead to a contract dying strictly before the contractual term. The numbers of contract of age k still alive at the beginning of period t is time dependent and denoted by $n_{k,t}$, with $k = 1, \ldots, K-1$. The probability that such a contract (i, k), $i = 1, \ldots, n_{kt}, k = 1, \ldots, K$ is closing at period t is:

$$P_t[Y_{i,k,t} = 1] = p_{k,t}, \tag{3.46}$$

where $p_{k,t}$ is the intensity for age k and date t.

The introduction of an unobservable stochastic time factor in an intensity model allows for differentiated effects of the factors depending on the age of the contract. For instance, for loans and a single risk factor, these effects are expected to be smaller at the beginning of the contract, or close to the contractual term. A stochastic intensity model with factor and logit

[6]Indeed, it is easily checked that the log-likelihood function is a sum $\sum_{k=1}^{K} L_k(y, p_k)$, say. The asymptotic independence follows from the expression of the information matrix, since the Hessian of the log-likelihood function becomes a diagonal matrix.

specification of the intensities is:

$$P[Y_{i,k,t} = 1 | F_t, \varepsilon_{k,t}] = [1 + \exp(\alpha_k + \beta_k F_t + \eta \varepsilon_{k,t})]^{-1}, \tag{3.47}$$

where F_t is the common factor and $\varepsilon_{k,t}$ are age-specific errors.

iii) The consequences of stochastic intensity

Without stochastic intensity effects $F_{k,t}$ and $\varepsilon_{k,t}$, and with constant α_k, the intensities:

$$P[Y_{i,k,t} = 1] = [1 + \exp \alpha]^{-1},$$

imply a lifetime following a geometric distribution. The introduction of stochastic variables in the intensities has two different effects. At the individual level, the marginal distribution of the lifetime is no longer geometric, but can feature negative duration dependence, that is, an intensity function decreasing with the age for instance. At the joint level, the presence of a common factor creates complicated patterns of dependence between the lifetimes of two individuals of a same cohort.

iv) Longevity risk

A first historical introduction of stochastic intensity with dynamic factor is due to Lee, Carter (1992) for the analysis of mortality in a given population. The Lee-Carter methodology described below is still the basic model used for life insurance and pension funds design and pricing.

Let us denote by $p_{k,t}$ the mortality intensity of an individual of age k at period t. This intensity is specified as:

$$\log\left(\frac{p_{k,t}}{1 - p_{k,t}}\right) = \alpha_k + \beta_k F_t, \quad k = 1, ..., K, \quad t = 1, ..., T, \tag{3.48}$$

where α_k and β_k are parameters and F_t is a stochastic factor, which corresponds to model (3.47) with $\eta = 0$ after a change of sign on α_k and β_k. The logit transform ensures that $p_{k,t}$ is between 0 and 1. [7] Since the unobserv-

[7] In their seminal paper, Lee and Carter used a log-transform which does not ensure this constraint.

able factor values are identifiable up to an affine transformation, we can set $E[F_t] = 0$ and $V[F_t] = 1$.

Then, Lee and Carter propose to use the published mortality tables which are providing a sample counterpart $\hat{p}_{k,t}$ of $p_{k,t}$, and to write the approximate factor model:

$$\log\left(\frac{\hat{p}_{k,t}}{1-\hat{p}_{k,t}}\right) = \alpha_k + \beta_k F_t + u_{k,t}, \tag{3.49}$$

where $u_{k,t}$ is a Gaussian error term. Then they estimate α_k by $\hat{\alpha}_k = \frac{1}{T}\sum_{t=1}^{T} \log\left(\frac{\hat{p}_{k,t}}{1-\hat{p}_{k,t}}\right)$, and deduce approximations of β_k and F_t by applying a singular value decomposition on the $T \times K$ matrix X with elements $X_{t,k} = \log\left(\frac{\hat{p}_{k,t}}{1-\hat{p}_{k,t}}\right) - \hat{\alpha}_k$. Specifically, the estimates of the factor values are given by vector $\hat{F} = (\hat{F}_1, ..., \hat{F}_T)'$, which is the eigenvector of the $T \times T$ matrix XX' associated with the largest eigenvalue and normalized such that $\hat{F}'\hat{F}/T = 1$. The estimates of the factor sensitivities are $\hat{\beta} = (\hat{\beta}_1, ..., \hat{\beta}_K)' = Y'\hat{F}/T$.

This methodology has been applied to the main developed countries using the data publicly available in the human mortality data base of Berkeley University and Max Planck Institute [8]. For instance, the results for France are summarized in Figures 3.7 and 3.8. The analysis is performed separately for female and male.

[Insert Figure 3.7: Estimated mortality factor values for French female and male, 1950-2007.]

[Insert Figure 3.8: Estimated intercepts and factor loadings for French female and male, 1950-2007.]

From Figure 3.7 we immediately observe that the mortality factor F_t features a (stochastic) downward trend. It corresponds to the general increase of human lifetime, that is, the average increase by about three months of residual lifetime every year. This increase varies in time and across genders. In Fig-

[8] Data can be downloaded from the web-page www.mortality.org.

ure 3.8 we display estimated intercepts α_k (left panels) and factor loadings β_k (right panels) for 1-year age classes $[k-1,k)$ with $k = 1, ..., 111$ years [9]. From the estimated intercepts, we see that the historical average mortality rate in period 1950-2007 features a non-monotonous pattern with respect to age. Mortality is higher for children and old people. Moreover, the average mortality is generally larger for male than for female, in age classes between 20 and 80 years. Factor loadings are positive and overall decreasing with respect to age, that is, the effect of the decreasing mortality trend is less pronounced for older people. The pattern of the factor loadings is similar across female and male, but factor loadings are generally larger for female than male. Thus, the impact of the decreasing mortality trend seems overall more important for female than male.

Let us now discuss the above estimation procedure in view of the general results presented in this chapter. The traditional Lee-Carter model (3.48) corresponds to a qualitative factor model as considered in Section 3.2, where the qualitative observations $Y_{i,k,t} \sim \mathcal{B}(1, p_{k,t})$ are the death events in the different age classes. However, the canonical factors $a_{k,t} = \log\left(\frac{p_{k,t}}{1-p_{k,t}}\right)$ for $k = 1, ..., K$ admit a degenerate dependence structure (see Remark 3.3). A non-degenerate dependence structure is obtained by adding age class specific mortality risks:

$$a_{k,t} = \alpha_k + \beta_k F_t + \eta \varepsilon_{k,t}, \tag{3.50}$$

where the shocks $\varepsilon_{k,t}$ are $IIN(0,1)$ across age classes and time dates, and $\eta > 0$ is the standard deviation of the class effects. Model (3.50) can be estimated by means of the CSA approach described in Section 3.3. Specifically, Theorem 3.3 implies that the CSA ML estimators are obtained from the spectral decomposition of the $K \times K$ matrix $X'X/T$. In particular, the estimates of the factor sensitivities correspond to the eigenvector associated with the largest eigenvalue of matrix $X'X/T$ (appropriately rescaled). In Appendix A.4, we show that the spectral decompositions of matrices XX' and $X'X$ are strongly related, namely, these matrices share the same non-zero

[9] The last age class includes people who are 110 years old or more.

eigenvalues and associated eigenvectors. Thus, the CSA ML estimates of the factor sensitivities coincide with those obtained from the standard procedure used in the literature. The CSA approach however also provides estimates for the standard deviation of the age class effects. In our empirical illustration with the mortality data in French in the period 1950-2007, the estimates are $\hat{\eta} = 0.089$ for females, and $\hat{\eta} = 0.097$ for males. They are small compared to the estimates of the common factor loadings. Hence, age class specific mortality effects do not seem very important for the considered datasets.

The basic estimation approach can be improved in several directions:

a) Even if the estimation method is close to the CSA approach with static factor, this factor is clearly dynamic with a trend. This type of extension will be considered in Chapter 4.

b) It is also possible to take into account the asymptotic variance of $\log\left(\frac{\hat{p}_{k,t}}{1-\hat{p}_{k,t}}\right)$, which depends on the level $p_{k,t}$ and the number of individuals of age k at date t (Population-at-Risk). Indeed, the size of the PaR is small for large ages and the information less accurate; this arises for the individuals, who are intuitively the most sensitive to longevity factors. Accounting for the asymptotic variance of the estimated canonical factors leads to VGA estimates of the model parameters [see Section 3.2 iv)].

c) Finally, several factors can be introduced. Typically the longevity factors are not necessarily the same for male and female, for workers or executives, for European and American.

3.5 Factor Analysis of Dependence

The general methodology can also be followed to understand the structure of dependence between two qualitative (or discrete) variables. Indeed, it is important to allow for different factors impacting the marginal distributions of two qualitative risks, or the dependence between these two risks. For this purpose, it is useful to introduce a suitable reparameterization of the joint distribution of two qualitative variables.

i) An appropriate parameterization for a 2×2 contingency table

Let us denote X and Z the two qualitative variables of interest, and $Y = (X, Z)$ the qualitative variable representing both of them. The alternatives for X [resp. Z, Y] are k, $k = 1, \ldots, K$ [resp. j, $j = 1, \ldots, J$; (k, j), $k = 1, \ldots, K$, $j = 1, \ldots, J$]. The distribution of Y at date t can be represented in a (K, J) contingency table, where:

$$P_t[Y_{i,t} = (k, j)] = p_{k,j,t}.$$

Table 3.1: (K, J) Contingency Table.

$X \quad Z$	$1 \ldots j \ldots J$
1	$\vdots$
$\vdots$	$\vdots$
k	$\cdots\ p_{k,j,t}\ \cdots$
$\vdots$	$\vdots$
K	$\vdots$

A reparameterization of the contingency table with real parameters, able to distinguish the marginal and dependence features, has been introduced in the eighties for the analysis of tendency surveys [see e.g. Koenig, Nerlove, Oudiz (1979), Nerlove (1983), Nerlove, Press (1986)]. It is called **log-linear probability model**. We use this parameterization to introduce the canonical factors. The idea is to separate in the log-probabilities the marginal effects of alternatives k and j from their cross-effects. More precisely, we consider the following decomposition:

$$\log p_{k,j,t} = \mu_t + a^1_{k,t} + a^2_{j,t} + a^{1,2}_{k,j,t}, \tag{3.51}$$

where:

$$\sum_{k=1}^{K} a^1_{k,t} = 0,\ \sum_{j=1}^{J} a^2_{j,t} = 0,\ \sum_{k=1}^{K} a^{1,2}_{k,j,t} = 0,\ \forall j,\ \sum_{j=1}^{J} a^{1,2}_{k,j,t} = 0,\ \forall k, \tag{3.52}$$

the leading term μ_t being deduced by the unit mass restriction.

Alternatively, model (3.51) can be seen as a special polytomous logit model:

$$p_{k,j,t} = \exp(a^1_{k,t} + a^2_{j,t} + a^{1,2}_{k,j,t}) / \left[\sum_{k,j} \exp(a^1_{k,t} + a^2_{j,t} + a^{1,2}_{k,j,t}) \right]. \tag{3.53}$$

ii) Factor analysis of a pair of dichotomous variables

For illustration, let us consider a pair of dichotomous variables and denote as usual in this framework their alternatives as 0, 1. By taking into account the restrictions in (3.52), the new parameters can be all written as functions of $a^1_{1,t} = a_{1,t}$ (say), $a^2_{1,t} = a_{2,t}$ (say), $a^{1,2}_{1,1,t} = a_{3,t}$ (say). We get:

$$\left\{ \begin{array}{lcl} \log p_{1,1,t} & = & \mu_t + a_{1,t} + a_{2,t} + a_{3,t}, \\ \log p_{1,0,t} & = & \mu_t + a_{1,t} - a_{2,t} - a_{3,t}, \\ \log p_{0,1,t} & = & \mu_t - a_{1,t} + a_{2,t} - a_{3,t}, \\ \log p_{0,0,t} & = & \mu_t - a_{1,t} - a_{2,t} + a_{3,t}, \end{array} \right. \tag{3.54}$$

and:

$$\left\{ \begin{array}{lcl} a_{1,t} & = & \frac{1}{4}[\log p_{1,1,t} + \log p_{1,0,t} - \log p_{0,1,t} - \log p_{0,0,t}], \\ a_{2,t} & = & \frac{1}{4}[\log p_{1,1,t} + \log p_{0,1,t} - \log p_{1,0,t} - \log p_{0,0,t}], \\ a_{3,t} & = & \frac{1}{4}[\log p_{11,t} + \log p_{0,0,t} - \log p_{1,0,t} - \log p_{0,1,t}]. \end{array} \right. \tag{3.55}$$

The canonical factors $a_{1,t}$ and $a_{2,t}$ have a positive impact on the probabilities of the events defined by $X = 1$, and $Z = 1$, respectively. To interpret the third canonical factor, note that $a_{3,t}$ can also be written as:

$$a_{3,t} = \frac{1}{4} \log \left(\frac{p_{1,1,t} p_{0,0,t}}{p_{1,0,t} p_{0,1,t}} \right). \qquad (3.56)$$

It takes value zero if and only if variables X and Z are independent at date t. It takes its maximal value $+\infty$, if either $p_{1,0,t}$, or $p_{0,1,t}$ is equal to zero. The minimal value $-\infty$ is achieved, when either $p_{1,1,t}$, or $p_{0,0,t}$, is 0. Thus, $a_{3,t}$ is a measure of the dependence between X and Z. By introducing different factors for $a_{1,t}$, $a_{2,t}$ and $a_{3,t}$, we can interpret these factors in terms of either marginal, or cross-effects.

iii) Illustration: The effect of the financial crisis

We illustrate the above methodology with an application to the analysis of the dynamics of the cross-sectional distribution of stock returns during the recent financial crisis. Let us consider the stocks that define the S&P 500 index. Let $r_{i,t}$, for $i = 1, ..., n$ and $n = 500$, and $r_{m,t}$ denote the return at day t of stock i, and the return of the index, respectively. Returns are percentage and concern the time period from May 7, 2007 to May 6, 2011. Figure 3.9 displays the time series of the S&P 500 index returns.

[Insert Figure 3.9: Time series of S&P 500 daily percentage returns, 2007/05/07 - 2011/05/06.]

We observe a period of large volatility between September 2008 and June 2009 corresponding to the recent financial crisis.

We consider the stock return $r_{i,t} - r_{m,t}$ of asset i in excess of the market, and discretize the support of this variable into four subsets, that are $I = \{r_{i,t} - r_{m,t} < -\lambda\}$, $II = \{-\lambda \leq r_{i,t} - r_{m,t} < 0\}$, $III = \{0 \leq r_{i,t} - r_{m,t} < \lambda\}$ and $IV = \{r_{i,t} - r_{m,t} \geq \lambda\}$, respectively, where $\lambda > 0$ is a threshold independent of asset and time. The threshold λ is fixed at $\lambda = 1.802$, which corresponds to the 75% quantile of $|r_{i,t} - r_{m,t}|$ across assets and dates in our sample. The subsets I, II, III and IV correspond to a large negative return, a moderate negative return, a moderate positive return and a large positive return, respectively. At each date t, we compute the cross-sectional frequency of stocks with returns in the subsets I-IV and study the dynamics of these frequencies.

The occurrence of a stock return in a subset I-IV can be characterized by means of two dichotomous variables. Let $X_{i,t} = 1$, if $r_{i,t} - r_{m,t} > 0$, and $= 0$, otherwise. Moreover, let $Z_{i,t} = 1$, if $|r_{i,t} - r_{m,t}| > \lambda$, and $= 0$, otherwise. Hence, $X_{i,t}$ is the indicator of a positive stock return and $Z_{i,t}$ is the indicator of a large absolute stock return. Then, subsets I-IV are characterized by $I = \{X_{i,t} = 0, Z_{i,t} = 1\}$, $II = \{X_{i,t} = 0, Z_{i,t} = 0\}$, $III = \{X_{i,t} = 1, Z_{i,t} = 0\}$ and $IV = \{X_{i,t} = 1, Z_{i,t} = 1\}$, respectively.

The evolution of the cross-sectional marginal distributions of the dichotomous variables X and Z are given in Figure 3.10.

[Insert Figure 3.10: Time series of probabilities $p_{1\cdot,t}$ and $p_{\cdot 1,t}$.]

The market return can be interpreted as a weighted average of individual returns. The value of $p_{1\cdot,t} = p_{11,t} + p_{12,t}$ gives information on the skewness of the cross-sectional distribution of individual returns. This marginal probability is equal to 0.5 (resp., larger, smaller than), if the median is equal to the mean (resp., larger, smaller than). From the first panel of Figure 3.10 (see also Table 3.1), we see that the distribution of X is in average moderately left skewed. However, we observe a large variability of this probability over time, and some periodic behaviour: periods in which there is a large number of assets performing better than the market are followed by periods in which much more assets underperform. The marginal probability $p_{\cdot 1,t} = p_{11,t} + p_{21,t}$ of Z is a market adjusted measure of individual risk. During the recent financial crisis, we get simultaneously an increased market volatility (see Figure 3.9), but also an increase of the market adjusted risks.

We provide in Figure 3.11 the evolutions of the contingency tables and in Figure 3.12 the evolutions of the log-linear parameters.

[Insert Figure 3.11: Time series of probabilities $p_{00,t}$, $p_{01,t}$, $p_{10,t}$ and $p_{11,t}$.]

[Insert Figure 3.12: Time series of factors $a_{1,t}$, $a_{2,t}$ and $a_{3,t}$.]

Factor $a_{2,t}$ is another measure of market adjusted risk, with an evolution largely similar to the evolution of $p_{\cdot 1,t}$. More interesting is the evolution of the cross-sectionl measure of dependence $a_{3,t}$ between X and Z. We might expect a kind of positive risk premium for market adjusted returns, that is extreme returns compensated by right skewed distribution, or equivalently more periods with positive values of factor $a_{3,t}$. This feature is clearly not observed on the third panel of Figure 3.12. This dependence measure is very erratic over time, as if we get a stochastic dependence, and this stochastic dependence fluctuates around 0 corresponding to the cross-sectional independence hypothesis.

3.6 Summary

The maximum likelihood method is complicated in factor models with unobservable factors. It can be approximated by the CSA and VGA approaches. In qualitative models with static factors, these approximated estimation methods are easy to implement, if (i) the models are written in terms of well-chosen canonical factors, and (ii) the canonical factors are linear functions of a reduced number of Gaussian underlying factors. The methodology can be applied to dichotomous or multinomial probit and logit models as well as to duration models or log-linear probability models with unobservable factors. The approach is especially relevant for longevity analysis.

R E F E R E N C E S

Anderson, T. (2003): *An Introduction to Multivariate Statistical Analysis*, Wiley, 3rd Edition.

Basel Committee on Banking Supervision (2001): "The New Basel Capital Accord", Consultative Document of the Bank for International Settlements, April 2001, Part 2: Pillar 1.

Basel Committee on Banking Supervision (2003): "The New Basel Capital Accord", Consultative Document of the Bank for International Settlements, April 2003, Part 3: The Second Pillar.

Berkson, J. (1944): "Application of the Logistic Function to Bioassay", JASA, 39, 357-365.

Black, F. (1976): "Studies of Stock Price Volatility Changes", Proceedings of the 1976 Meeting of the American Statistical Association, Business and Economics Statistics Section, 177-181.

Blake, D., Cairns, A., and K., David (2006): "A Two Factor Model for Stochastic Mortality with Parameter Uncertainty: Theory and Calibration", Journal of Risk and Insurance, 73, 687-718.

Gouriéroux, C., and A., Monfort (1989): "Statistics and Econometric Models", Cambridge University Press.

Gouriéroux, C., and A., Monfort (1996): "Simulation Based Estimation Methods", Oxford University Press.

Gouriéroux, C., and A., Monfort (2010): "Granularity in Qualitative Factor Model", Journal of Credit Risk, 5, Winter 2009/10.

Jennrich, R., and D., Thayer (1977): "A Note on Lawley's Formulas for Standard Errors in Maximum Likelihood Factor Analysis", Psychometrika, 38, 571-580.

Koenig, H., Nerlove, M., and G., Oudiz (1979): "Modèles log-linéaires pour l'analyse des données qualitatives", Annales de l'INSEE, 36, 31-83.

Lawley, D., and A., Maxwell (1971): *Factor Analysis as a Statistical Method*, Elsevier, New-York, 2nd edition.

Lee, R., and L., Carter (1972): "Modeling and Forecasting US Mortality", Journal of the American Statistical Association, 87, 659-675.

McFadden, D. (1973): "Conditional Analysis of Qualitative Choice Behavior", in *Frontiers of Econometrics*, P. Zarembka, ed., 105-142, New-York, Academic.

McFadden, D. (1976): "Quantal Choice Analysis: A Survey", Annals of Economic and Social Measurement, 5, 363-390.

Merton, R. (1974): "On the Pricing of Corporate Debt: The Risk Structure of Interest Rates", Journal of Finance, 29, 449-470.

Nerlove, M. (1983): "Expectations, Plans and Realizations in Theory and Practice", Econometrica, 51, 1251-1279.

Nerlove, M., and S., Press (1986): "Multivariate Log-linear Probability Models", in *Advances in Statistical Analysis and Statistical Computing: Theory and Application*, R. Mariano, ed., 117-174, Greenwich, JAI Press.

Vasicek, O. (1991): "Limiting Loan Loss Probability Distribution", DP KMV Corporation.

3.7 Appendix: CSA Maximum Likelihood Estimator in Factor Model

(*) Inverse and determinant of matrix Ω

Let us denote by $\tilde{\beta}_l = \beta_l/(\beta_l'\beta_l)^{1/2}$ the column vectors of matrix β rescaled to have a unit norm. Set $\tilde{\beta}_l$, $l = 1, \ldots, L$, can be completed in order to get a set $\tilde{\beta}_l$, $l = 1, \ldots, KS$, which forms an orthonormal basis of $I\!R^{SK}$. We get:

$$\Omega = \sum_{l=1}^{L} \beta_l\beta_l' + \eta^2 Id = \sum_{l=1}^{L}(\eta^2 + \beta_l'\beta_l)\tilde{\beta}_l\tilde{\beta}_l' + \eta^2 \sum_{l=L+1}^{KS} \tilde{\beta}_l\tilde{\beta}_l'.$$

This provides the spectral decomposition of matrix Ω. We deduce that:

$$\det\Omega = (\eta^2)^{KS-L}\prod_{l=1}^{L}(\eta^2 + \beta_l'\beta_l),$$

$$\Omega^{-1} = \sum_{l=1}^{L} \frac{1}{\eta^2 + \beta_l'\beta_l}\tilde{\beta}_l\tilde{\beta}_l' + \sum_{l=L+1}^{KS} \frac{1}{\eta^2}\tilde{\beta}_l\tilde{\beta}_l'$$

$$= -\sum_{l=1}^{L} \beta_l\beta_l'\frac{1}{\eta^2(\eta^2 + \beta_l'\beta_l)} + \frac{1}{\eta^2}Id.$$

(**) The CSA log-likelihood function

We have:

$$\frac{1}{T}L^{\text{CSA}}(\theta) \propto -\frac{1}{2}\log\det\Omega(\beta,\eta^2) - \frac{1}{2T}\sum_{t=1}^{T}(\hat{a}_t - \alpha)'\Omega(\beta,\eta^2)^{-1}(\hat{a}_t - \alpha).$$

(***) CSA estimator of α

The first-order condition with respect to α is:

$$\sum_{t=1}^{T}[\Omega(\beta,\eta^2)]^{-1}(\hat{a}_t-\alpha)=0$$

$$\iff \sum_{t=1}^{T}(\hat{a}_t-\alpha)=0$$

$$\iff \hat{\alpha}_T=\frac{1}{T}\sum_{t=1}^{T}\hat{a}_t=\bar{a}_T.$$

(****) Concentrated CSA log-likelihood function

Therefore, the log-likelihood function concentrated with respect to α is:

$$\begin{aligned}\frac{1}{T}\tilde{L}^{\mathrm{CSA}}(\beta,\eta^2) &\propto -\frac{1}{2}\log\det\Omega(\beta,\eta^2)-\frac{1}{2T}\sum_{t=1}^{T}[(\hat{a}_t-\hat{\alpha}_T)'\Omega(\beta,\eta^2)^{-1}(\hat{a}_t-\hat{\alpha}_T)]\\ &= -\frac{1}{2}\log\det\Omega(\beta,\eta^2)-\frac{1}{2}Tr[\Omega(\beta,\eta^2)^{-1}\hat{V}_T],\end{aligned}$$

where $\hat{V}_T=\frac{1}{T}\sum_{t=1}^{T}(\hat{a}_t-\bar{a}_T)(\hat{a}_t-\bar{a}_T)'$ is the historical variance-covariance matrix of the estimated canonical factors, and Tr denotes the trace operator which computes the sum of the diagonal elements of a square matrix. From the expressions of $\det\Omega$ and Ω^{-1} derived in (*), we deduce:

$$\begin{aligned}\frac{1}{T}\tilde{L}^{\mathrm{CSA}}(\beta,\eta^2) &\propto -\frac{KS-L}{2}\log\eta^2-\frac{1}{2}\sum_{l=1}^{L}\log(\eta^2+\beta_l'\beta_l)-\frac{1}{2\eta^2}Tr(\hat{V}_T)\\ &\quad+\frac{1}{2}\sum_{l=1}^{L}\frac{\beta_l'\hat{V}_T\beta_l}{\eta^2(\eta^2+\beta_l'\beta_l)}.\end{aligned}$$

(*****) Estimators of β and η^2

Let us consider the first-order condition with respect to β_l without taking into account the orthogonality restrictions between the sensitivity vectors.

We get:

$$\frac{1}{T}\frac{\partial \tilde{L}^{\text{CSA}}}{\partial \beta_l} = \left\{ -\frac{1}{\eta^2+\beta_l'\beta_l} - \frac{\beta_l'\hat{V}_T\beta_l}{\eta^2(\eta^2+\beta_l'\beta_l)^2} \right\} \beta_l + \frac{\hat{V}_T\beta_l}{\eta^2(\eta^2+\beta_l'\beta_l)} = 0.$$

This first-order condition implies that $\hat{V}_T\beta_l$ and β_l are proportional, that is, β_l is an eigenvector of matrix $\hat{V}_T$.

Let us denote by $\hat{e}_l$ an eigenvector of $\hat{V}_T$ with unit norm proportional to β_l, and by $\hat{\lambda}_l$ the associated eigenvalue. We have:

$$\beta_l = \gamma_l^{1/2}\hat{e}_l,$$

where $\gamma_l = \beta_l'\beta_l$. By substituting in the first-order condition, we get an equation which defines γ_l :

$$-\frac{1}{\eta^2+\gamma_l} - \frac{\hat{\lambda}_l\gamma_l}{\eta^2(\eta^2+\gamma_l)^2} + \frac{\hat{\lambda}_l}{\eta^2(\eta^2+\gamma_l)} = 0$$

$$\iff \quad \gamma_l = \beta_l'\beta_l = \hat{\lambda}_l - \eta^2.$$

Let us finally concentrate with respect to the optimal $\beta_l's$. We get:

$$\begin{aligned}
&\frac{1}{T}\tilde{L}^{\text{CSA}}(\eta^2) \\
\propto \quad & -\frac{KS-L}{2}\log\eta^2 - \frac{1}{2}\sum_{l=1}^{L}\log\hat{\lambda}_l - \frac{1}{2\eta^2}Tr(\hat{V}_T) + \frac{1}{2}\sum_{l=1}^{L}\frac{(\hat{\lambda}_l-\eta^2)}{\eta^2} \\
= \quad & -\frac{KS-L}{2}\log\eta^2 - \frac{1}{2\eta^2}\left[Tr(\hat{V}_T) - \sum_{l=1}^{L}\hat{\lambda}_l\right] - \frac{1}{2}\sum_{l=1}^{L}\log\hat{\lambda}_l - \frac{L}{2}.
\end{aligned}$$

The first-order condition with respect to η^2 provides :

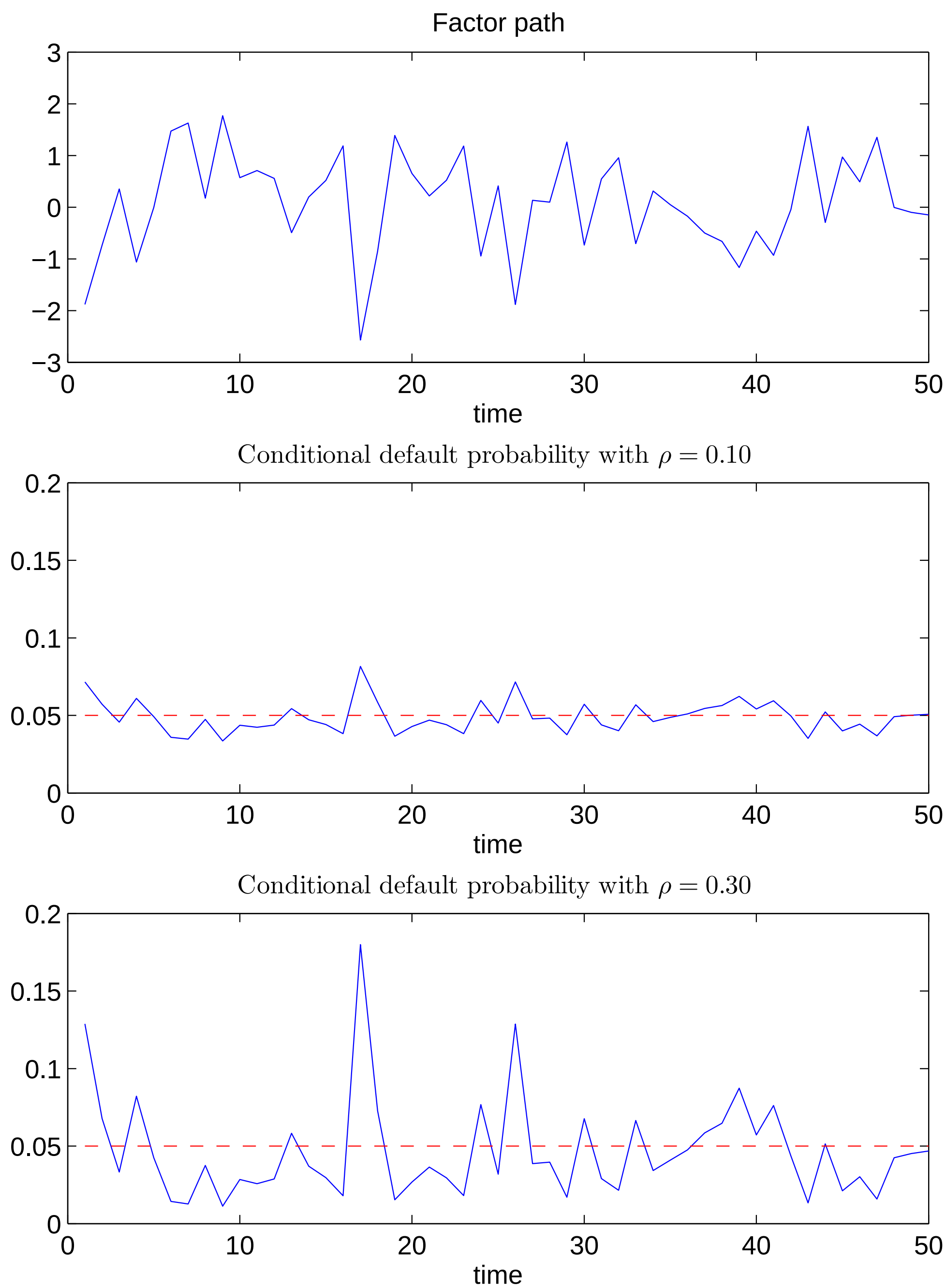
$$\eta^2 = \frac{Tr(\hat{V}_T) - \sum_{l=1}^{L}\hat{\lambda}_l}{KS-L}.$$

Since $Tr(\hat{V}_T) = \sum_{l=1}^{KS} \hat{\lambda}_l$, the corresponding value of the concentrated CSA log-likelihood is equal (up to an additive constant) to:

$$-\frac{KS-L}{2} \log \left(\sum_{l=L+1}^{KS} \hat{\lambda}_l \right) - \frac{1}{2} \sum_{l=1}^{L} \log \hat{\lambda}_l.$$

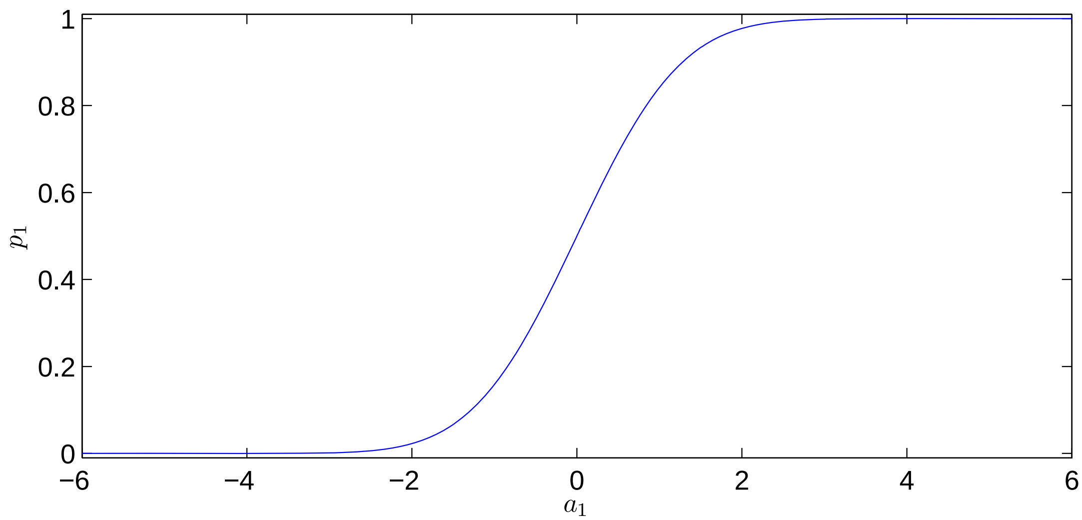
This value is maximized when the L largest eigenvalues are selected. This proves Proposition 3.3.

Figure 3.1: Time-varying conditional default probability.



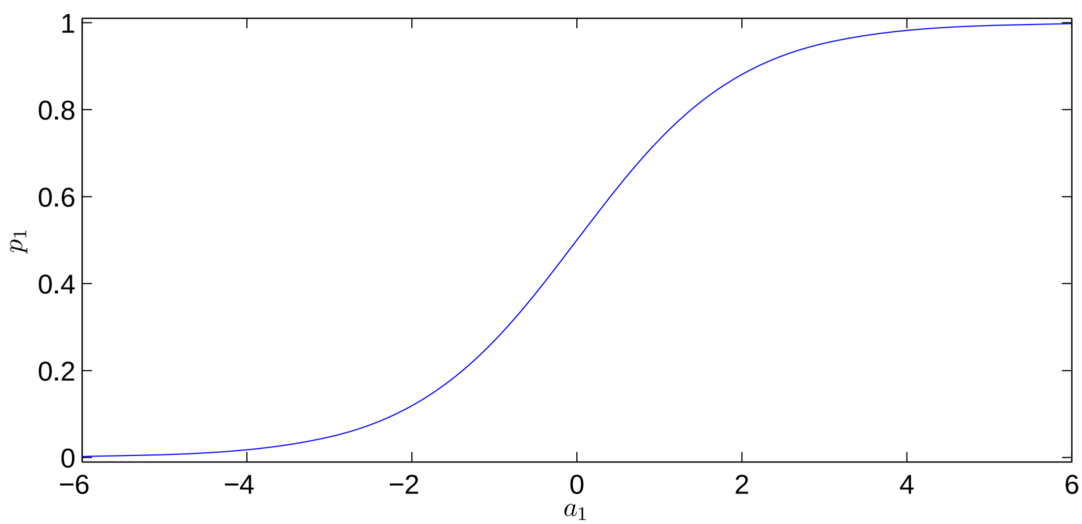
The upper panel displays a simulated path of the factor $F_t \sim IIN(0,1)$ of time length 50 periods. The middle panel displays the corresponding path of conditional default probability (solid line) in a cohort of firms with unconditional default probability $PD = 0.05$ (dashed horizontal line) and asset correlation $\rho = 0.10$. The lower panel displays the pattern of conditional default probability with asset correlation $\rho = 0.30$.

Figure 3.2: Parameterization of the dichotomous probit model.



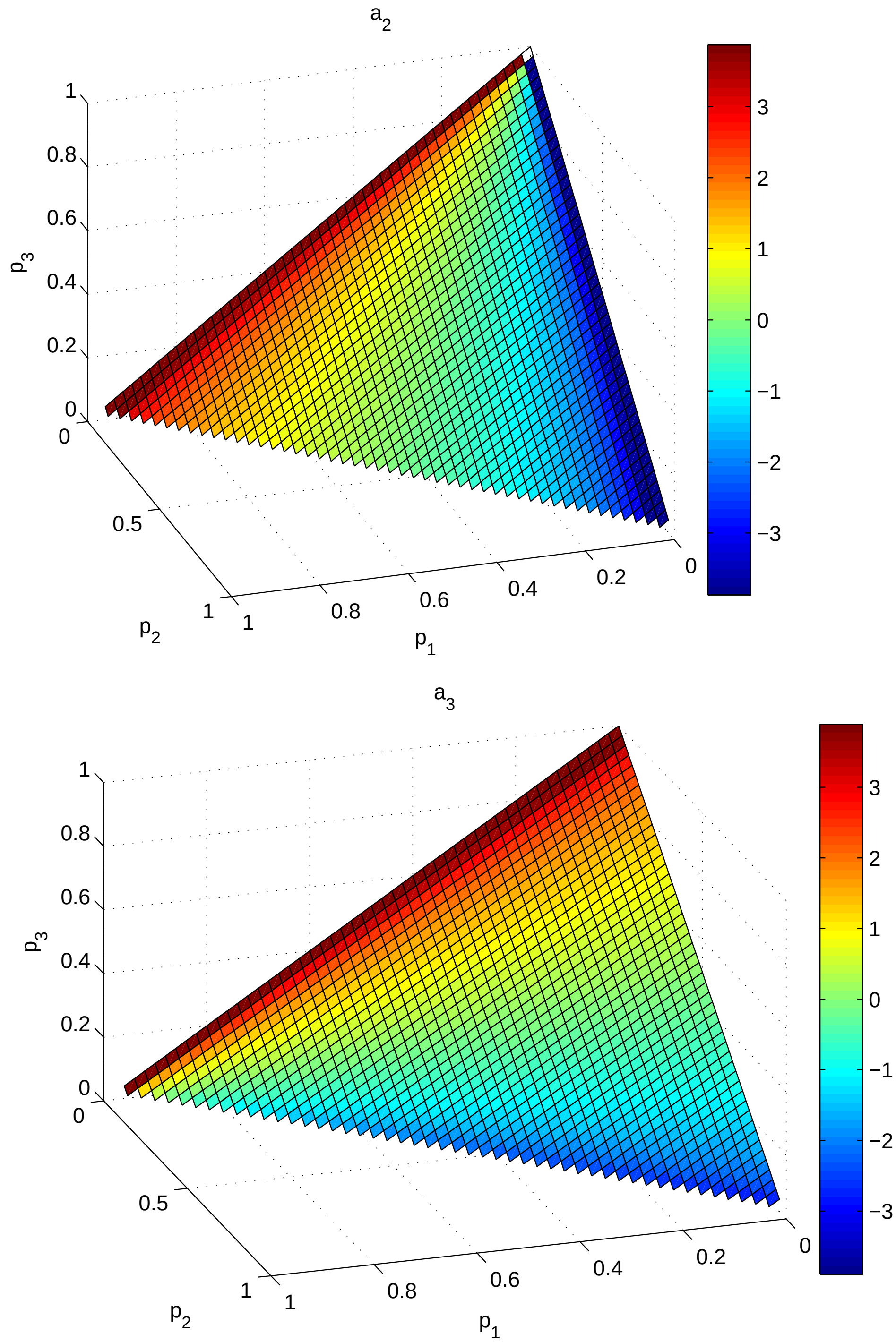
The figure displays the probability p_1 as a function of the canonical factor a_1 in the dichotomous probit model.

Figure 3.3: Parameterization of the dichotomous logit model.



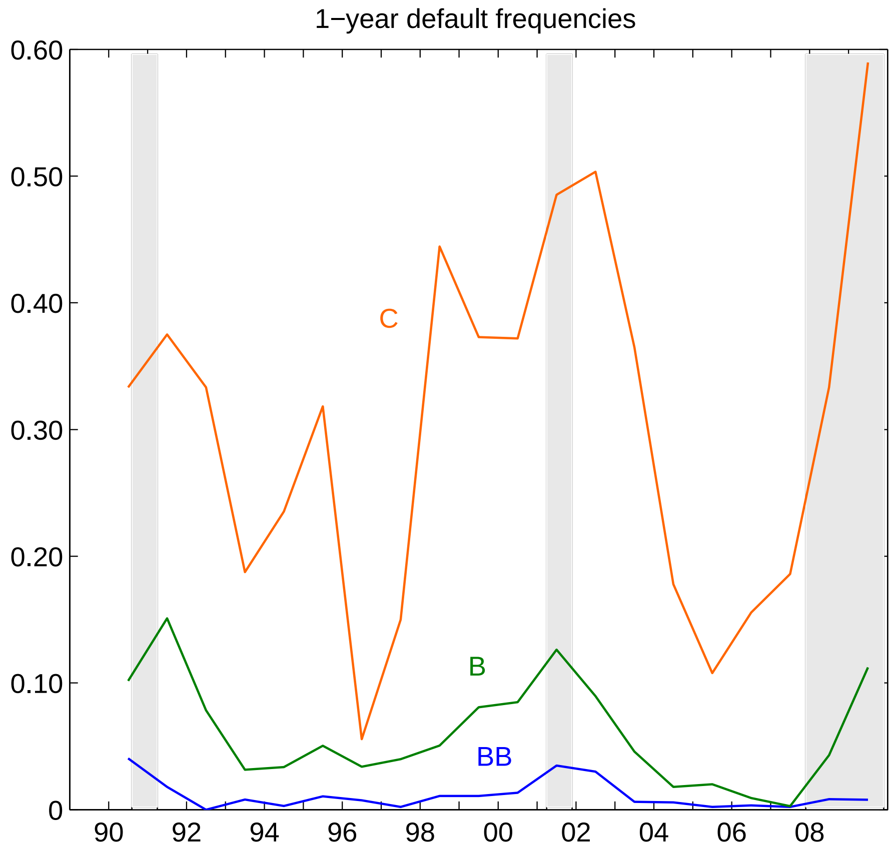
The figure displays the probability p_1 as a function of the canonical factor a_1 in the dichotomous logit model.

Figure 3.4: Canonical factors in the multinomial logit model.



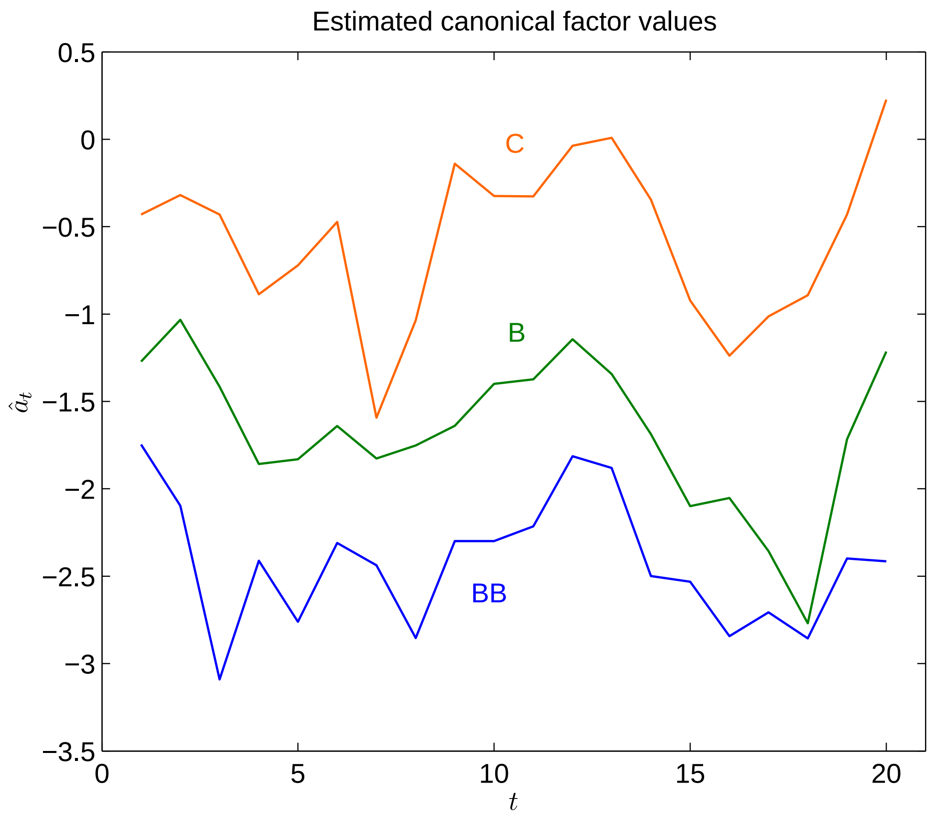
The upper panel displays the level-color map of the canonical factor a_2 as a function of the elementary probabilities (p_1, p_2, p_3) in the multinomial logit model with $J = 3$ alternatives. Colors on the simplex correspond to function values. The lower panel displays the level-color map for the canonical factor a_3.

Figure 3.5: S&P US corporate default frequencies for rating classes BB, B and C.



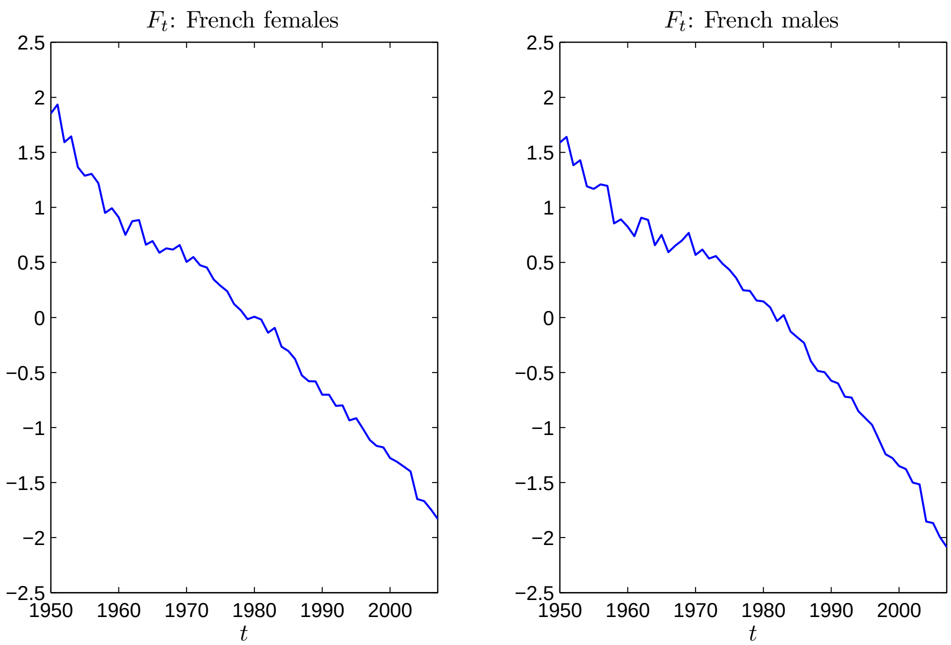
The Figure displays the series of S&P US corporate default frequencies for rating classes BB, B, C in the period 1990-2009. Shaded periods correspond to NBER recessions in US.

Figure 3.6: Estimated canonical factor values for rating classes BB, B and C.



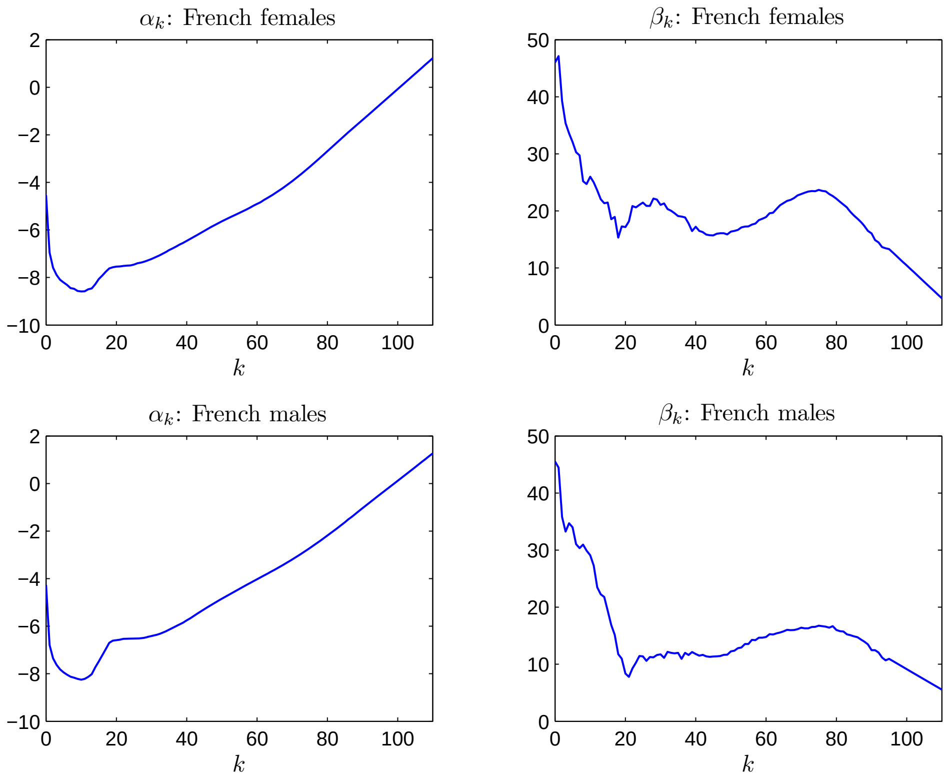
The Figure displays the series of estimated canonical factor values $\hat{a}_{k,t}$ for rating classes BB, B, C in the period 1990-2009.

Figure 3.7: Estimated mortality factor values for French female and male, 1950-2007.



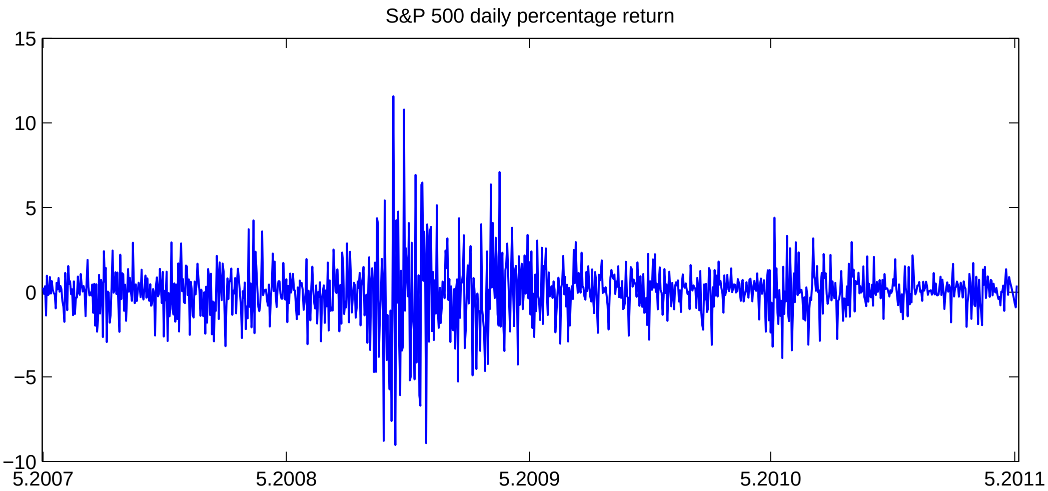
The Figure displays the series of estimated mortality factor values for French female (left) and male (right) in the period 1950-2007. The factor is normalized such that its historical mean is zero and its historical variance is 1.

Figure 3.8: Estimated intercepts and factor loadings for French female and male, 1950-2007.



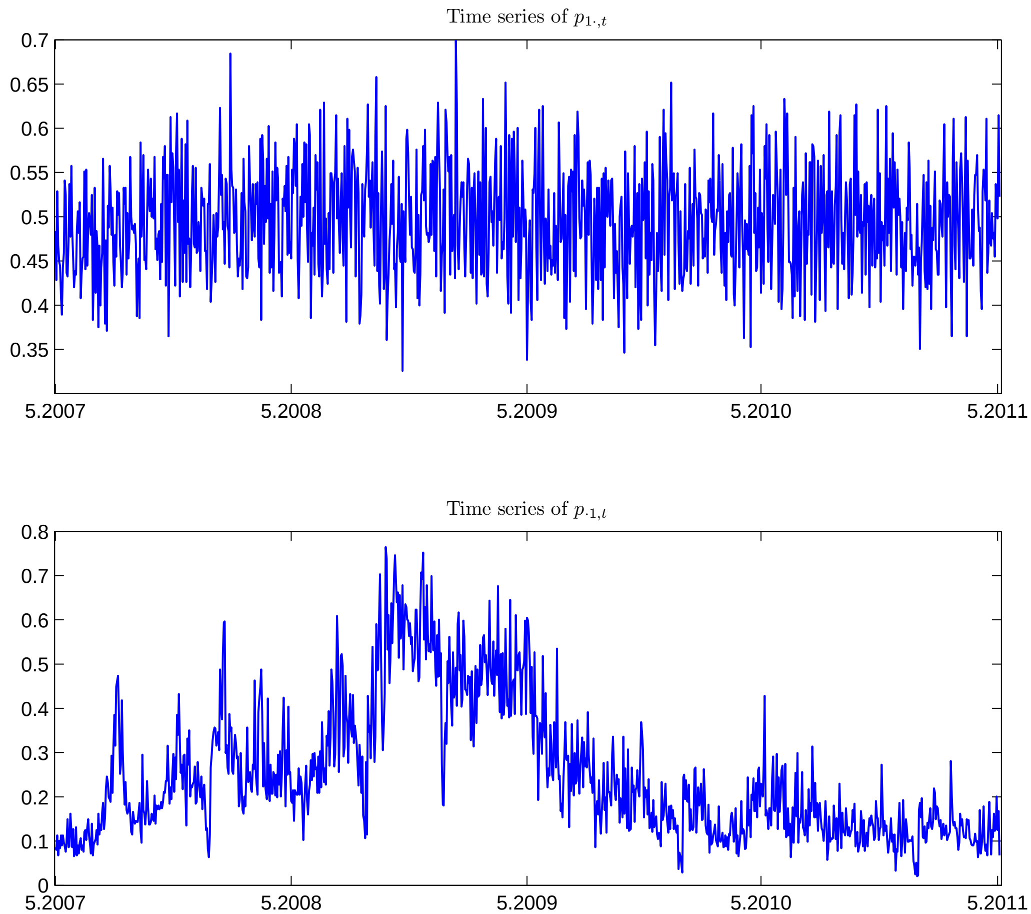
The Figure displays the estimated intercepts α_k and factor loadings β_k for French female and male in 1-year age classes $[k-1, k)$, for $k = 1, ..., 111$ years. The factor is normalized such that its historical mean is zero and its historical variance is 1.

Figure 3.9: Time series of S&P 500 daily percentage returns, 2007/05/07 - 2011/05/06.



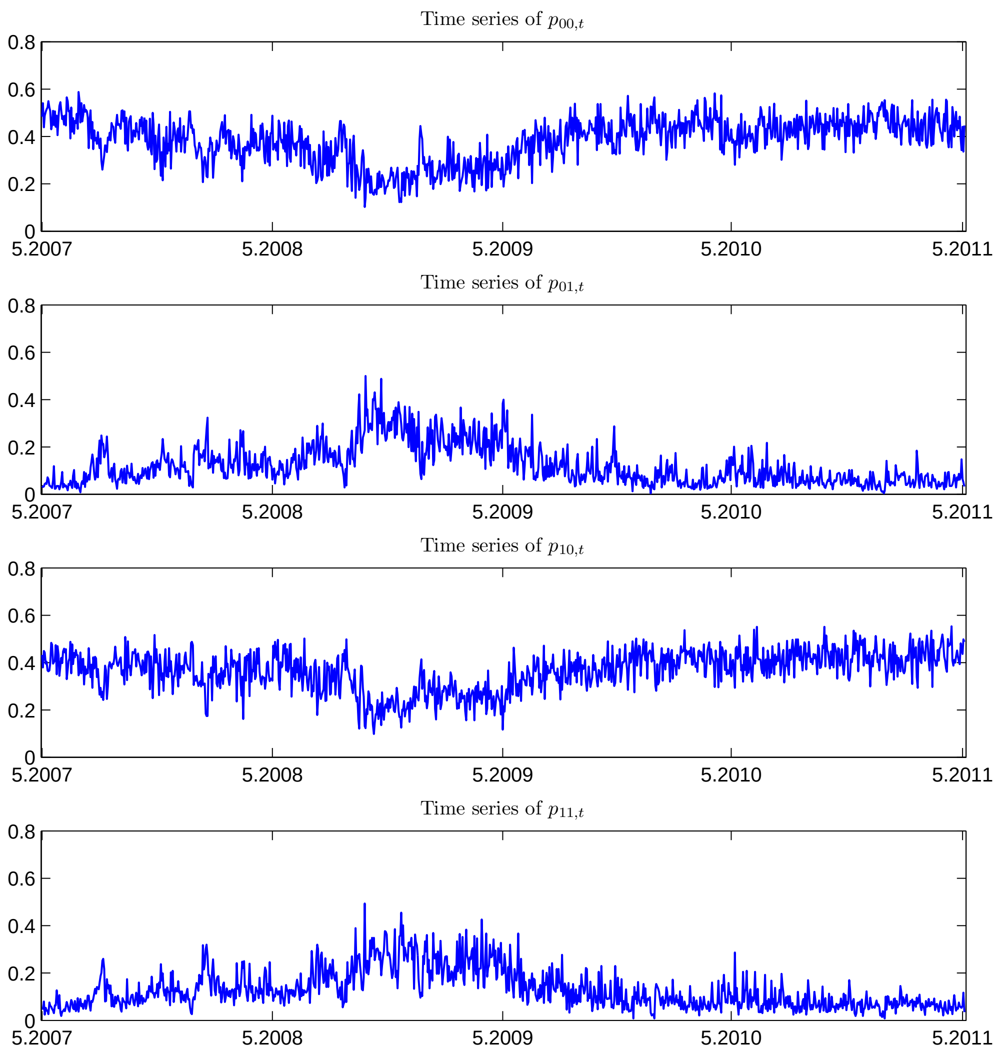
The Figure displays the time series of daily percentage returns of the S&P 500 index in the period from 2007/05/07 to 2011/05/06.

Figure 3.10: Time series of probabilities $p_{1\cdot,t}$ and $p_{\cdot 1,t}$.



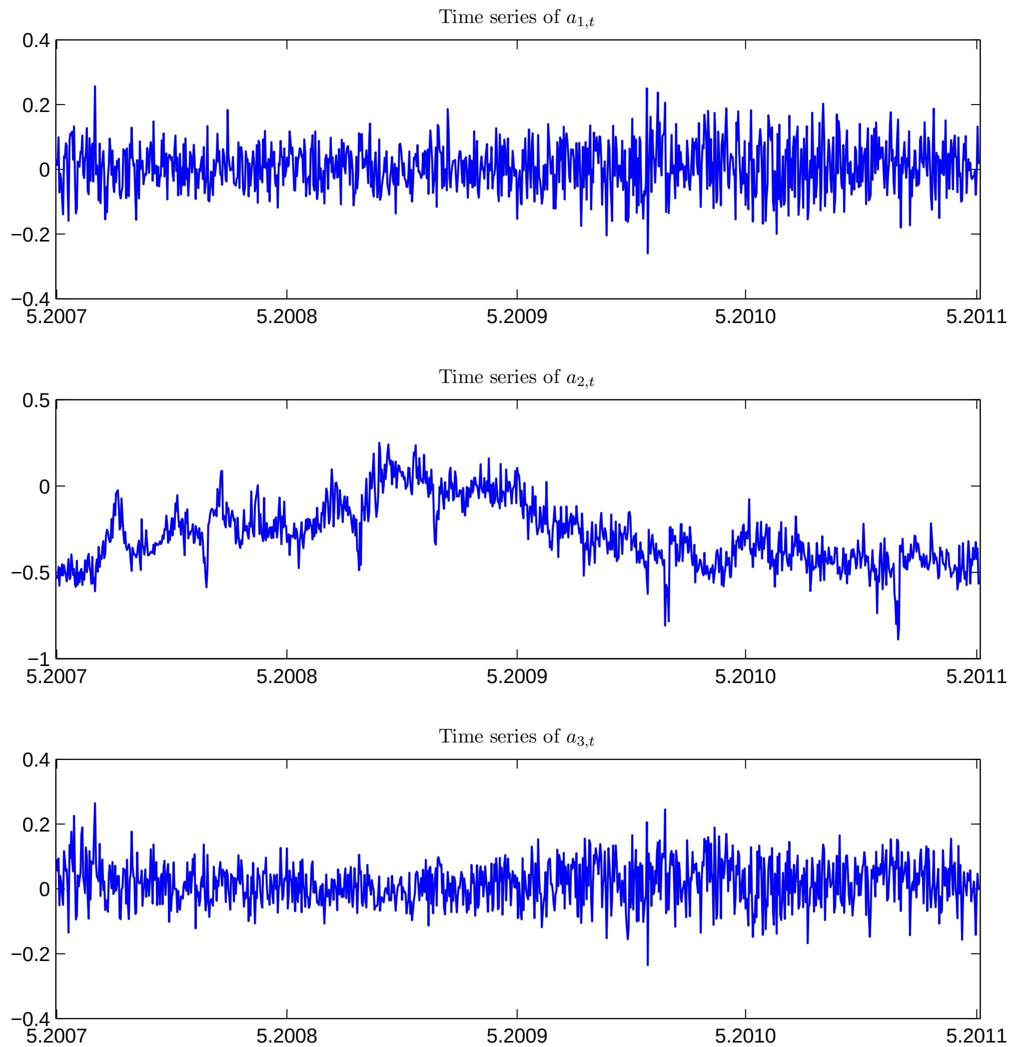
The upper panel displays the time series of probability $p_{1\cdot,t} = P_t[X_{i,t} = 1]$ of positive stock return. The lower panel displays the time series of probability $p_{\cdot 1,t} = P_t[Z_{i,t} = 1]$ of large absolute stock return. Stock returns are in excess of the market return.

Figure 3.11: Time series of probabilities $p_{00,t}$, $p_{01,t}$, $p_{10,t}$ and $p_{11,t}$.



This Figure displays the time series of probabilities $p_{00,t}$ (first panel), $p_{01,t}$ (second panel), $p_{10,t}$ (third panel), and $p_{11,t}$ (fourth panel), where $p_{kl,t} = P_t[X_{i,t} = k, Z_{i,t} = l]$ for $k, l = 0, 1$.

Figure 3.12: Time series of factors $a_{1,t}$, $a_{2,t}$ and $a_{3,t}$.



This Figure displays the time series of factors $a_{1,t}$ (upper panel), $a_{2,t}$ (middle panel) and $a_{3,t}$ (bottom panel).

Table 3.1: Sample 2×2 contingency table of variables $X_{i,t}$ an $Z_{i,t}$.

		$X_{i,t}$ 0	$X_{i,t}$ 1	
$Z_{i,t}$	0	0.384	0.366	0.75
	1	0.121	0.129	0.25
		0.505	0.495	

Table 3.2: The 2×2 contingency table of variables $X_{i,t}$ an $Z_{i,t}$ under the independence assumption.

		$X_{i,t}$ 0	$X_{i,t}$ 1	
$Z_{i,t}$	0	0.379	0.371	0.75
	1	0.126	0.124	0.25
		0.505	0.495	

Chapter 4

Nonlinear Dynamic Panel Model

The application of granularity theory to estimation is presented in this chapter for general nonlinear dynamic panel models with common factors. These models can feature nonlinear dynamics in both the measurement and state equations. Intuitively, the specification distinguishes between the dynamics at the individual level through the lagged individual observations (micro-dynamics), and the dynamics at the aggregate level through the factors (macro-dynamics). Consequently, the parameterization of these models involves macro-parameters as well as micro-parameters.

In Section 4.1, we explain why the GA methodology remains simple in qualitative models with dynamic Gaussian latent factors. Indeed, in these specifications with macro-parameters only, both the CSA and GA approximated models are linear state space models, for which the standard Kalman filter applies. The results for general models with both macro- and micro-parameters and nonlinear factor dynamics are described in Section 4.2. We explain how to derive estimators of the micro- and macro-parameters, which are asymptotically efficient when both the cross-sectional dimension n and the time dimension T tend to infinity. We also provide approximations of the factor values (see also Chapter 5). As expected, the rates of convergence differ: they are $1/\sqrt{nT}$ for the micro-parameters, $1/\sqrt{T}$ for the macro-parameters and $1/\sqrt{n}$ for the factor values, respectively. A sketch of the proof of the asymptotic results is given in Section 4.3, where we also introduce the CSA and GA maximum likelihood estimators of the parameters. The application to stochastic migration models is presented in Section 4.4. These models are used for a joint analysis of the corporate rating migrations in an homogeneous set of companies. An empirical analysis using S&P rating data of US companies in the period 1990-2009 is presented in Section 4.5.

4.1 Qualitative Model with Gaussian Dynamic Factor

i) The model

The static qualitative factor model of Section 3.2 can be extended to include factor dynamics. Let us assume individual qualitative observations such that:

$$P[Y_{i,k,t} = j | a_t] = p(j; a_{k,t}), \quad j = 1, \ldots, J, \quad k = 1, \ldots, K, \quad t = 1, \ldots, T, \tag{4.1}$$

where $a_t = (a'_{1,t}, \ldots, a'_{K,t})' \in I\!R^{KS}$ denotes the canonical factor. Moreover, suppose that the canonical factors are noisy linear transformations of a smaller number $L < KS$ of underlying macro-factors $F_t \in I\!R^L$ with a Gaussian Vector Autoregressive (VAR) dynamic:

$$a_t = \alpha + \beta F_t + \eta w_t, \tag{4.2}$$

where:

$$F_t = \Phi F_{t-1} + \varepsilon_t, \tag{4.3}$$

and the errors (w_t), (ε_t) are independent, such that $w_t \sim IIN(0, Id)$ and $\varepsilon_t \sim IIN(0, \Omega)$, say. Model (4.1)-(4.3) above is a state space model with static nonlinear measurement equations (4.1) and Gaussian dynamic linear state equation (4.2)-(4.3).

ii) Approximated linear state space model

For large cross-sectional dimensions n_k of the cohorts, the nonlinear state space model can be approximated by a Gaussian linear state space model, for which standard softwares based on the linear Kalman filter are available. [1] These softwares can be used for parameter estimation as well as prediction of the future individual qualitative variables, or filtering of the unobservable factor values.

Let us consider the fixed effect maximum likelihood estimator of the canonical factor values [see Section 3.2 ii)]:

$$\hat{a}_{k,t} = \arg\max_{a_{k,t}} \sum_{j=1}^{J} n_{j,k,t} \log p(j; a_{k,t}). \tag{4.4}$$

Under identification conditions, we know that asymptotically [see (3.20)-(3.21)]:

$$\hat{a}_t \overset{d}{\sim} N\left(a_t, \hat{\Sigma}_{n,t}\right), \tag{4.5}$$

where $\hat{\Sigma}_{n,t} = \text{diag}[\hat{\Sigma}_{k,t}/n_k]$, and that the $\hat{a}_t$, t varying, are (asymptotically) independent. Thus, the nonlinear static measurement equations written on individual qualitative observation $y_{i,k,t}$, $i = 1, \ldots, n_k$, $k = 1, \ldots, K$, $t = 1, \ldots, T$ can be asymptotically replaced by the linear measurement equations (4.5) written on the aggregate statistics $\hat{a}_t$, $t = 1, \ldots, T$. In other words, the initial model can be replaced by the following VGA linear state space model:

State equation:

$$F_t = \Phi F_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim IIN(0, \Omega); \tag{4.6}$$

VGA measurement equation:

[1] See e.g. the *sspace* object in EVIEWS, or the Kalman function in the *Control and System Toolbox* in MATLAB.

$$\hat{a}_t = \alpha + \beta F_t + u_t, \quad u_t \sim IIN(0, \eta^2 Id + \hat{\Sigma}_{n,t}). \tag{4.7}$$

The GA appears by means of the additional variance-covariance matrix $\hat{\Sigma}_{n,t}$ in the measurement equation. If $n_k = \infty$, for any $k = 1, \ldots, K$, this term would disappear. This yields the CSA linear state space model:

State equation:

$$F_t = \Phi F_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim IIN(0, \Omega); \tag{4.8}$$

CSA measurement equation:

$$\hat{a}_t = \alpha + \beta F_t + u_t, \quad u_t \sim IIN(0, \eta^2 Id). \tag{4.9}$$

Thus, both the CSA and VGA approximated models are linear state space models and can be analyzed by the standard linear Kalman filter (see the Review Appendix A.5). The Kalman filter is used to estimate parameters α, β, μ, Φ, Ω and η, and to filter the latent factor values.

4.2 Asymptotically Efficient Estimators

i) The model

Let us now consider the general nonlinear dynamic model with unobservable factor (see Chapter 1, Section 1.3). For expository purpose, the model is presented for a single cohort. It is defined by its transition densities which are parameterized as follows:

State equation: The conditional density of f_t given f_{t-1} is $g(f_t|f_{t-1};\theta)$.

Measurement equations: The conditional density of $y_{i,t}$ given $y_{i,t-1}$ and f_t is $h(y_{i,t}|y_{i,t-1}, f_t; \beta)$.

Conditional on the factor path, the individual histories $(y_{i,t})$, $i = 1, \cdots, n$, are independent Markov processes, with a same transition density $h(y_{i,t}|y_{i,t-1}, f_t; \beta)$ between $t-1$ and t depending on the factor value f_t. The factor varies stochastically in time according to a Markov process with transition density $g(f_t|f_{t-1};\theta)$. The model involves a vector of micro-parameters β that characterize the dynamics at the individual level (micro-dynamics), as well as a vector of macro-parameters θ that characterize the dynamics of the factor (macro-dynamics). The unknown true values of these parameters are denoted β_0 and θ_0, respectively. When the unobservable stochastic factors (f_t) are integrated out, the model for the observable variables features both cross-sectional dependence and non-Markovian serial dependence.

If the variables $y_{i,t}$, $i = 1, \dots, n$, and f_t were observable at each date, the joint density (conditional on the initial observations) would be:

$$l^*(\underline{y_T}, \underline{f_T}; \beta, \theta) = [\prod_{t=1}^{T} g(f_t|f_{t-1};\theta)] \prod_{t=1}^{T} \prod_{i=1}^{n} h(y_{it}|y_{i,t-1}, f_t; \beta). \tag{4.10}$$

Thus, the latent log-likelihood function could be decomposed as:

$$\begin{aligned} \mathcal{L}^*(\underline{y_T}, \underline{f_T}; \beta, \theta) &= \log l^*(\underline{y_T}, \underline{f_T}; \beta, \theta) \\ &= \mathcal{L}^M(\underline{f_T}; \theta) + \sum_{t=1}^{T} \mathcal{L}^{CS}(y_t|y_{t-1}, f_t; \beta), \end{aligned} \tag{4.11}$$

where:

$$\mathcal{L}^M(\underline{f_T}; \theta) = \sum_{t=1}^{T} \log g(f_t|f_{t-1};\theta), \tag{4.12}$$

is the log-likelihood corresponding to the macro-economic factor, called the **latent macro log-likelihood function**, and:

$$\mathcal{L}^{CS}(y_t|y_{t-1}, f_t; \beta) = \sum_{i=1}^{n} \log h(y_{i,t}|y_{i,t-1}, f_t; \beta), \tag{4.13}$$

is the log-likelihood corresponding to individual transitions between dates $t-1$ and t. It is called the **latent cross-sectional micro log-likelihood function**.

The different log-likelihood functions described in (4.11)-(4.13) are latent, since they assume the latent factors observable. As already mentioned, the true log-likelihood function is deduced by integrating out the unobservable factors. It is given by

$$\log l(\underline{y_T}; \beta, \theta),$$

where:

$$l(\underline{y_T}; \beta, \theta) = \int \dots \int [\prod_{t=1}^{T} g(f_t|f_{t-1}; \theta)] \prod_{t=1}^{T} \prod_{i=1}^{n} h(y_{i,t}|y_{i,t-1}, f_t; \beta) \prod_{t=1}^{T} df_t. \tag{4.14}$$

This log-likelihood function has a complicated expression, which involves a multiple integral with a huge dimension equal to T times the number of factors. In particular, the dimension of this integral tends to infinity with time dimension T.

When n and T are large, it is possible to derive asymptotically efficient estimators of both types of parameters without having to compute the huge integral in (4.14). This is shown next.

ii) The estimation method

If micro-parameter β were known, the factor value at date t could be approximated by the **fixed effects estimator:**

$$\begin{aligned} \hat{f}_{nt}(\beta) &= \arg\max_{f_t} \mathcal{L}^{CS}(y_t|y_{t-1}, f_t; \beta) \\ &= \arg\max_{f_t} \sum_{i=1}^{n} \log h(y_{i,t}|y_{i,t-1}, f_t; \beta). \end{aligned} \tag{4.15}$$

The name fixed effects is used because estimator $\hat{f}_{nt}(\beta)$ is computed by treating f_t as a parameter in the latent cross-sectional micro-likelihood, that is, by considering the factor values as fixed time effects.

However, micro-parameter β is unknown, and thus the factor approximations $\hat{f}_{nt}(\beta)$ as well. But these values can be reintroduced in the latent micro-likelihood functions aggregated over time, to get a function of the observations $\underline{y_T}$ and parameter β only. This leads to an estimator of β defined by:

$$\hat{\beta}_{nT} = \arg\max_{\beta} \sum_{t=1}^{T} \sum_{i=1}^{n} \log h[y_{i,t}|y_{i,t-1}, \hat{f}_{nt}(\beta); \beta]. \tag{4.16}$$

Equivalently, it can also be derived by considering the solution in β in the joint optimization problem:

$$\max_{\beta, f_1, \ldots, f_T} \sum_{t=1}^{T} \sum_{i=1}^{n} \log h(y_{i,t}|y_{i,t-1}, f_t; \beta). \tag{4.17}$$

The definition of the estimator through (4.17) shows that the unknown factor values have been treated as nuisance parameters. Such an approach might create an **incidental parameter** problem, since the number of nuisance parameters tends to infinity with T [Neyman and Scott (1948)]. This problem does not exist in our framework, where the cross-sectional dimension is much larger than the time dimension (see Proposition 4.1 below and the discussion thereafter).

The estimator of the micro-parameter can be introduced in the expression of the fixed effects estimator of the factor value to get an approximation of factor value at date t:

$$\hat{f}_{nT,t} = \hat{f}_{n,t}(\hat{\beta}_{nT}). \tag{4.18}$$

These approximated factor values can serve as proxies for the unobserved factor values. This leads to the following estimator of the macro-parameter θ:

$$\hat{\theta}_{nT} = \arg\max_{\theta} \sum_{t=1}^{T} \log g(\hat{f}_{nT,t}|\hat{f}_{nT,t-1}; \theta). \tag{4.19}$$

The estimator $\hat{\theta}_{nT}$ maximizes the latent macro log-likelihood $\mathcal{L}^M$ after replacing the factor values by their proxies.

Remark 4.1: The models presented in Chapters 2, 3, and in Section 4.1 were models with macro-parameters only. In such cases, the fixed effects estimator of the first step is a function of the individual observations only, that is,

$$\hat{f}_{nT,t} = \hat{f}_{n,t} = \arg\max_{f_t} \sum_{i=1}^{n} \log h(y_{i,t}|y_{i,t-1}, f_t). \tag{4.20}$$

iii) Asymptotic properties of the estimators

The asymptotic properties of the estimators of micro- and macro-parameters introduced above have been derived in Gagliardini, Gouriéroux (2009) (see also the discussion in Section 4.3). Their asymptotic distribution involves information matrices corresponding to the latent macro-likelihood and cross-sectional micro-likelihood. More precisely:

i) The cross-sectional information matrix at date t is:

$$I^{CS}(t) = E_0 \left[-\frac{\partial^2 \log h(y_{i,t}|y_{i,t-1}, f_t; \beta_0)}{\partial(\beta', f')'\partial(\beta', f')} | \underline{f_t} \right]. \tag{4.21}$$

It involves the conditional expectation of the second-order derivative matrix of the micro log-density w.r.t. the micro-parameter and the factor value, given the current and past history of factor values $\underline{f_t} = (f_t, f_{t-1}, \cdots)$. This information matrix can be written in block form as follows:

$$I^{CS}(t) = \begin{bmatrix} I_{\beta\beta}(t) & I_{\beta f}(t) \\ & \\ I_{f\beta}(t) & I_{ff}(t) \end{bmatrix}. \tag{4.22}$$

ii) The macro information matrix is:

$$I^{M} = E_0 \left[-\frac{\partial^2 \log g(f_t|f_{t-1}; \theta_0)}{\partial\theta\partial\theta'} \right]. \tag{4.23}$$

We have the following proposition valid under the set of regularity conditions in Gagliardini, Gouriéroux (2009):

Proposition 4.1: *If the dimensions n, T tend to infinity such that $T^b/n = O(1)$, for $b > 1$, then:*

i) *The estimators are consistent and asymptotically normal:*

$$\begin{bmatrix} \sqrt{nT}(\hat{\beta}_{n,T} - \beta_0) \\ \\ \sqrt{T}(\hat{\theta}_{n,T} - \theta_0) \end{bmatrix} \xrightarrow{d} N\left(\begin{bmatrix} 0 \\ \\ 0 \end{bmatrix}, \begin{bmatrix} (I^*_{\beta\beta})^{-1} & 0 \\ \\ 0 & (I^M)^{-1} \end{bmatrix} \right),$$

where:

$$I^*_{\beta\beta} = E_0[I_{\beta\beta}(t) - I_{\beta f}(t) I_{ff}(t)^{-1} I_{f\beta}(t)].$$

ii) *The estimators are asymptotically efficient.*

iii) *For any date t, conditional on the factor path we have:*

$$\sqrt{n}(\hat{f}_{nT,t} - f_t) \xrightarrow{d} N[0, I_{ff}(t)^{-1}].$$

All estimators converge to their corresponding true values when both n and T tend to infinity at suitable relative rates (namely when n is infinitely larger than T in the limit). However, the convergence rates of the estimators differ: they are equal to $1/\sqrt{nT}$ for the micro-parameters, $1/\sqrt{T}$ for the macro-parameters and $1/\sqrt{n}$ for the factor values, respectively. Estimators $\hat{\beta}_{n,T}$ and $\hat{\theta}_{n,T}$ are asymptotically independent. Thus, asymptotically the inference on β and θ can be done separately. In other words, parameters β (resp. θ) are actually micro-parameters (resp. macro-parameters), since they do not include macro-information (resp. micro-information).

The estimators $\hat{\beta}_{nT}$ and $\hat{\theta}_{nT}$ are asymptotically efficient in the sense that they are asymptotically equivalent to the maximum likelihood estimators that maximize the true likelihood function (4.14). Intuitively, when both n and T are large, estimators $\hat{\beta}_{nT}$ and $\hat{\theta}_{nT}$ have the lowest possible variance within a very large class of regular consistent estimators. Let us now discuss the expressions of the asymptotic variance-covariance matrices of the estimators. The information matrix I^M corresponds to the Fisher information on θ when the factor values are observable. Thus, for large n, the replacement of these values $f_t, t = 1, \ldots, T$, by their approximations $\hat{f}_{nT,t}, t = 1, \ldots, T$, has no effect on the estimator of macro-parameter θ. This is because the approximation errors of order $O_p(1/\sqrt{n})$ on the factor values are irrelevant for estimation of parameter θ at rate $1/\sqrt{T}$, when $T/n \to 0$.

The information matrix $I^*_{\beta\beta}$ is the information matrix for β in the micro-model with parameters β, $f_1, \dots, f_T$. It does not coincide with $E_0[I_{\beta\beta}(t)]$, since the estimation errors on the factor values have to be taken into account for estimation of the micro-parameters at rate $1/\sqrt{nT}$. However, we observe that the matrix $I^*_{\beta\beta}$ does not depend on the selected dynamic factor model. We deduce that $\hat{\beta}_{n,T}$ is both asymptotically efficient and semi-parametrically efficient [see Gagliardini, Gouriéroux (2009) and the Review Appendix A.2].

Finally, the usual panel literature emphasizes the role of incidental parameters, that is, the fact that in some models the number of unknown parameters increases with sample size [see Lancaster (2000) for the discussion of incidental parameters in panel models with individual effects]. In our panel model with common factor, the incidental parameters are the factor values, whose number increases with the time dimension T. If the cross-sectional dimension n were fixed, the presence of incidental parameters would imply the inconsistency of $\hat{\beta}_{n,T}$ even for large T. By assuming that n also tends to infinity faster than T, not only the convergence, but also the asymptotic efficiency, are obtained.

4.3 Likelihood Expansions, CSA and GA Maximum Likelihood Estimators

The asymptotic properties of the estimators of micro- and macro-parameters presented in Section 4.2 rely on asymptotic expansions of the complicated likelihood function of the model. We describe below the principle of this expansion [see Gagliardini, Gouriéroux (2009) for complete proofs], since the same principle will also be used for prediction purpose (see Chapter 5). Moreover, these asymptotic expansions of the likelihood function are the basis for deriving the CSA and GA maximum likelihood estimators.

i) First-order expansion of the log-likelihood function

The joint density of the observations is [see (4.14)]:

$$
\begin{aligned}
l(\underline{y_T};\beta,\theta) &= \int \dots \int \prod_{t=1}^{T} \prod_{i=1}^{n} h(y_{i,t}|y_{i,t-1},f_t;\beta) \prod_{t=1}^{T} g(f_t|f_{t-1};\theta) \prod_{t=1}^{T} df_t \\
&= \int \dots \int \exp\left\{ \sum_{t=1}^{T} \sum_{i=1}^{n} \log h(y_{i,t}|y_{i,t-1},f_t;\beta) \right\} \prod_{t=1}^{T} g(f_t|f_{t-1};\theta) \prod_{t=1}^{T} df_t.
\end{aligned}
$$

For large n the integral with respect to the factor values can be approximated by expanding the integrand around its maximum w.r.t. the factor along the lines of the **Laplace approximation** [see Jensen (1995), Arellano, Bonhomme (2009), and Appendix 4.7 i)]. We get the following expansion:

Proposition 4.2: *If n, T tend to infinity, with $T^b/n = O(1)$, for $b > 1$, we have:*

$$
\begin{aligned}
\mathcal{L}_{nT}(\beta,\theta) &= \frac{1}{nT} \log l(\underline{y_T};\beta,\theta) \\
&= \mathcal{L}^*_{nT}(\beta) + \frac{1}{n} \mathcal{L}_{1,nT}(\beta,\theta) + o_p(1/n), \qquad (4.24)
\end{aligned}
$$

where:

$$
\mathcal{L}^*_{nT}(\beta) = \frac{1}{nT} \sum_{t=1}^{T} \sum_{i=1}^{n} \log h[y_{i,t}|y_{i,t-1},\hat{f}_{nt}(\beta);\beta],
$$

$$
\mathcal{L}_{1,nT}(\beta,\theta) = -\frac{1}{2}\frac{1}{T} \sum_{t=1}^{T} \log \det I_{nt}(\beta) + \frac{1}{T} \sum_{t=1}^{T} \log g[\hat{f}_{nt}(\beta)|\hat{f}_{n,t-1}(\beta);\theta],
$$

and:

$$
I_{nt}(\beta) = -\frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2 \log h}{\partial f_t \partial f_t'} (y_{i,t}|y_{i,t-1},\hat{f}_{nt}(\beta);\beta).
$$

The decomposition above explains the main asymptotic results given in Proposition 4.1. Let us first consider the micro-parameters. They are involved in both components of the right hand side of decomposition (4.24).

However, since the second component is negligible w.r.t. the first one when n is large, the ML estimator of the micro-parameters is equivalent to the estimator based on the optimization of the micro log-likelihood $\mathcal{L}^*_{nT}$. This is exactly the definition of the fixed effects estimator $\hat{\beta}_{n,T}$ given in Section 4.2.

Let us now consider macro-parameters θ. They are involved in the second term of the expansion (4.24). Since the estimators of the micro-parameters converge faster than the rate $1/\sqrt{T}$, the maximum likelihood estimator of the macro-parameters can be approximated by the solutions of:

$$\max_\theta \mathcal{L}_{1,n,T}(\hat{\beta}_{n,T}, \theta)$$

$$\iff \max_\theta \sum_{t=1}^{T} \log g[\hat{f}_{nt}(\hat{\beta}_{n,T})|\hat{f}_{n,t-1}(\hat{\beta}_{n,T}); \theta].$$

This yields the estimator $\hat{\theta}_{n,T}$ introduced in Section 4.2. The asymptotic independence between the estimators of the micro- and macro-parameters is due to the additive decomposition of the log-likelihood function in (4.24), where the first component concerns β and the second one θ (since β can be replaced asymptotically by $\hat{\beta}_{n,T}$ in the second component).

In fact, the estimators $(\hat{\beta}_{nT}, \hat{\theta}_{nT})$ are asymptotically equivalent to the estimators derived by optimizing the first-order expansion of the log-likelihood function in Proposition 4.2. Let us denote by:

$$\mathcal{L}^{CSA}_{n,T}(\beta, \theta) = \mathcal{L}^*_{nT}(\beta) + \frac{1}{n}\mathcal{L}_{1,n,T}(\beta, \theta), \tag{4.25}$$

the **cross-sectional asymptotic (CSA) log-likelihood function**, and define the **CSA ML estimators** as:

$$(\hat{\beta}^{CSA}_{nT}, \hat{\theta}^{CSA}_{nT}) = \arg\max_{\beta,\theta} \mathcal{L}^{CSA}_{n,T}(\beta, \theta). \tag{4.26}$$

The CSA ML estimators are asymptotically equivalent to the estimators $(\hat{\beta}_{nT}, \hat{\theta}_{nT})$, and in particular asymptotically efficient.

ii) Granularity adjustment

The true maximum likelihood estimators of the parameters can be approximated more accurately by considering a second-order expansion of the log-likelihood function w.r.t $1/n$. We have:

$$\mathcal{L}_{nT}(\beta,\theta) = \mathcal{L}^*_{nT}(\beta) + \frac{1}{n}\mathcal{L}_{1,n,T}(\beta,\theta) + \frac{1}{n^2}\mathcal{L}_{2,n,T}(\beta,\theta) + o_p(1/n^2), \quad (4.27)$$

where the additional term $\mathcal{L}_{2,n,T}$ has a closed form expression [see Gagliardini, Gouriéroux (2009)] and does not involve integrals w.r.t. the unobservable factors. This second-order expansion defines the **GA log-likelihood function**:

$$\mathcal{L}^{GA}_{n,T}(\beta,\theta) = \mathcal{L}^*_{n,T}(\beta) + \frac{1}{n}\mathcal{L}_{1,n,T}(\beta,\theta) + \frac{1}{n^2}\mathcal{L}_{2,n,T}(\beta,\theta). \quad (4.28)$$

Then, granularity adjusted estimators are defined by maximizing the GA log-likelihood function:

$$(\hat{\beta}^{GA}_{nT}, \hat{\theta}^{GA}_{nT}) = \arg\max_{\beta,\theta} \mathcal{L}^{GA}_{n,T}(\beta,\theta). \quad (4.29)$$

In the general framework, the granularity adjustment is more important for the estimators of the macro-parameters, whose speed of convergence is slower. The granularity adjustment allows to modify the bias at order $1/n$ of the CSA estimator $\hat{\theta}^{CSA}_{nT}$. In particular, it is possible to show that the difference between $\hat{\theta}^{GA}_{nT}$ and the true ML estimator of θ is of order $o_p(1/n)$, while this difference would be of order $O_p(1/n)$ for the CSA estimator $\hat{\theta}^{CSA}_{nT}$ and for the estimator $\hat{\theta}_{nT}$ defined in Section 4.2. Thus, the GA and the true ML estimator of the macro-parameters are equivalent at order $1/n$.

iii) Newton-Raphson algorithm

It is easily checked that estimators asymptotically equivalent to the GA estimators are obtained by applying a single iteration in an appropriate Newton-Raphson algorithm with the CSA estimator as starting value [see

Appendix 4.7 ii)]. We have:

$$\begin{pmatrix} \hat{\beta}_{nT}^{GA} \\ \hat{\theta}_{nT}^{GA} \end{pmatrix} = \begin{pmatrix} \hat{\beta}_{nT}^{CSA} \\ \hat{\theta}_{nT}^{CSA} \end{pmatrix} + \left[-\frac{\partial^2 \mathcal{L}_{nT}^{CSA}(\hat{\beta}_{nT}^{CSA}, \hat{\theta}_{nT}^{CSA})}{\partial(\beta', \theta')' \partial(\beta', \theta')} \right]^{-1} \cdot \frac{\partial \mathcal{L}_{nT}^{GA}}{\partial(\beta', \theta')'} (\hat{\beta}_{nT}^{CSA}, \hat{\theta}_{nT}^{CSA}), \quad (4.30)$$

up to order $o_p(1/n^2)$ for the micro-parameters, and $o_p(1/n)$ for the macro-parameters, respectively.

4.4 Stochastic Migration Model

The Basel 2 regulation was not only asking for an accurate analysis of default risk and default correlation (see Section 3.1), but also of the risk associated with possible rating downgrades and upgrades [BCBS (2001),(2003)]. Indeed, the current rating has a significant impact on the value of the debt, and this effect has to be taken into account when assessing the risk of a credit portfolio. For this purpose, a dynamic analysis of the qualitative rating histories is required, with special focus on rating migration correlation. Unobservable dynamic factors are typically introduced in the models to create downgrade (resp. upgrade) correlation. Following the demand by regulators, stochastic migration models have been recently introduced in the academic literature, with special emphasis on corporate ratings and business cycle [see e.g. Gordy, Heitfield (2002), Gagliardini, Gouriéroux (2005a, b), Feng, Gouriéroux, Jasiak (2008)].

i) Stochastic transition matrices

Let us consider an homogenous subpopulation and individual qualitative histories $(y_{i,t}, t = 1, \ldots, T)$, for $i = 1, \ldots, n$. Variables $y_{i,t}$ are polytomous qualitative with K possible alternatives, denoted $k = 1, \ldots, K$. In the application to corporate bonds, these alternatives are the possible ratings, e.g. AAA, AA, A, BBB, ..., D in the Standard & Poor's (S&P) rating system.

Their number is typically either 8, or 10, depending whether ratings CCC, CC, C are put together, or not. However, the stochastic migration model described below can be applied to other frameworks as well.

As before, the dynamic model is defined by state and measurement equations.

State equation: The transition density of the factor is $g(f_t|f_{t-1};\theta)$;

Measurement equations: They are defined by the transition probabilities:

$$P[y_{i,t} = k|y_{i,t-1} = l, f_t; \beta] = \pi_{lk}(f_t; \beta), \text{ say,}$$

for $k, l = 1, \ldots, K$.

Since the individual observations are qualitative, the transition pdf of $y_{i,t}$ given $y_{i,t-1}$ and f_t is characterized by the (K, K) transition matrix:

$$\Pi(f_t; \beta) = [\pi_{lk}(f_t; \beta)]. \tag{4.31}$$

This transition matrix has nonnegative elements, which sum up to one by row. Its diagonal elements provide the probabilities to keep the same rating between dates $t-1$ and t, whereas the out-of-diagonal elements are the probabilities to migrate up or down, from one rating class to another one.

For a given factor history, the individual qualitative rating histories are independent, identically distributed. Each individual rating history is a Markov chain, which is time heterogenous since the transition matrices evolve in time. When the factor is considered stochastic, we get Markov chains with stochastic transition matrices. This justifies the alternative names given to this type of model, that are, **stochastic migration model**, **stochastic transition model**, or **model with stochastic intensity** (see also Section 3.4 for the two-state case with an absorbing state). When the factor is integrated out, the individual histories become dependent and the Markov property is lost. Intuitively, the whole past of all series is informative and needed to reconstitute approximately the current factor value.

There exist different migration models according to the specification of the transition matrix. Intuitively, each row of the transition matrix defines a probability distribution for which an ordered polytomous model (see Example 3.4 in Section 3.3), or a multinomial logit model (see Example 3.3 in Section 3.3), can be chosen. Moreover, the models for the different rows can be linked as seen below for the model usually considered for the analysis of rating histories.

ii) The dynamic ordered qualitative model for rating histories

This model is the direct extension of the SRF model of Section 3.1 to more than two alternatives and to a dynamic framework. In the spirit of Merton's structural model, the rating is based on the level of the log asset-to-liability ratio. More precisely, let us introduce a partition of the real line: $c_0 = -\infty < c_1 < c_2 < ... < c_{K-1} < c_K = +\infty$. We assume that:

$$y_{i,t} = k, \text{ if and only if } \; c_{k-1} < \log(A_{i,t}/L_{i,t}) \leq c_k, \tag{4.32}$$

for $k = 1, \ldots, K$, where $A_{i,t}$ and $L_{i,t}$ denote the asset value and the debt of firm i at date t. In this way, the rating classes are numbered in order of increasing credit quality, with alternative $k = 1$ typically corresponding to default [2]. Then, we have to define the conditional distribution of the log asset-to-liability ratio given the factor and the past individual histories. We assume that this dependence is through the most recent rating only. The latent model, which extends (3.3), is:

$$\log A_{i,t} - \log L_{i,t} = a_l + b_l F_t + \sigma_l u_{i,t}, \tag{4.33}$$

for companies with rating $y_{i,t-1} = l$ at date $t-1$, where $u_{i,t} \sim N(0,1)$ is independent of F_t. By comparing with equation (3.3), we see that the conditioning with respect to the last rating is equivalent to the creation at each

[2] In the basic Merton's model, $c_1 = 0$. In practice the constraint $c_1 = 0$ is generally not introduced, especially due to the regulatory definition of default. For instance, a default can be reported if the lender thinks that a failure of a company is highly probable in the next future, even if this failure has not yet occurred. Thus, threshold c_1 can be strictly positive and its magnitude depends on the more or less severe judgement of the lender.

date t of a set of K homogenous subpopulations, in which the corporations are grouped according to their previous rating.

Under (4.32)-(4.33) the stochastic transition probabilities are given by:

$$\pi_{l,k}(f_t;\beta) = \Phi\left(\frac{c_k - a_l - b_l f_t}{\sigma_l}\right) - \Phi\left(\frac{c_{k-1} - a_l - b_l f_t}{\sigma_l}\right), \tag{4.34}$$

where Φ is the cdf of the standard normal. These transition probabilities involve two types of micro-parameters, that are the parameters a_l, b_l, σ_l, $l = 1, \dots, K$ of the latent model for the individual asset-to-liability ratios, and the thresholds c_k, $k = 1, \dots, K-1$, used to define the ratings. Whereas a_l, b_l and σ_l appear in row l of the transition matrix only, the threshold parameters are in all rows, introducing links between rows. Moreover, while we have focused above on a model with a single factor in analogy to Section 3.1, the extension to include a multivariate factor F_t of dimension d, say, is straightforward. In this case, for each row l of the transition matrix we would have a $(d, 1)$ vector b_l of sensitivities to the different factors.

Finally, the measurement equations (4.34) are usually completed by a state equation corresponding to a Gaussian VAR model:

$$F_t = \mu + AF_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim IIN(0, \Omega), \tag{4.35}$$

where A is a (d, d) matrix of autoregressive coefficients.

iii) Identification

Parameter identification has to be considered carefully before applying any estimation method. The vector of model parameters is identified if it is not possible to find two distinct parameter vectors that imply the same distribution for the observable variables, that is, for the joint history of the ratings of the n firms [see e.g. Gourieroux, Monfort (1995)]. In the ordered qualitative model for ratings described in Section ii), the lack of identification can result from the loss in information incurred when passing from the quantitative scoring variable (the asset-to-liability ratio) to the qualitative rating, and from the unobservable factors being defined up to an invertible linear transformation. The identification restrictions have to account for the links existing between the rows of the transition matrix and have to conserve the distinction between micro- and macro-parameters.

For instance, in a stochastic migration model for rating histories with a single factor and absorbing state $k = K$, identification restrictions are $c_1 = a_1 = 0$ and $b_1 = \sigma_1 = 1$. Therefore, the identifiable micro-parameters are $\beta = (a_k, b_k, \sigma_k, c_k,\ k = 2, ..., K-1)$, whereas the macro-parameters are $\theta = (\mu, A, \Omega)$, that are all the parameters of the state equation. In the general case, we denote by β and θ the vectors of identifiable micro- and macro-parameters of the model, once suitable identification restrictions are imposed.

iv) Asymptotically efficient estimators

The asymptotically efficient estimators of the micro- and macro-parameters, and the factor approximations, are derived from the general results in Section 4.2, equations (4.16), (4.18) and (4.19). For expository purpose, we focus on a single-factor model.

Let us denote by $N_{k,t}$ and $N_{l,k,t}$ the number of companies in rating class k at date t, and the number of companies migrating from class l to class k between dates $t-1$ and t, respectively. The transition frequencies $\hat{\pi}_{l,k,t} = N_{l,k,t}/N_{l,t-1}$, for $l, k = 1, \dots, K$, are the empirical counterpart of the stochastic transition probabilities between dates $t-1$ and t. Then, the cross-sectional micro-density of the model at date t is given by:

$$\begin{aligned}
\mathcal{L}^{CS}(y_t|y_{t-1}, f_t; \beta) &= \sum_{l=1}^{K}\sum_{k=1}^{K} N_{l,k,t} \log \pi_{l,k,t}(f_t;\beta) \\
&= \sum_{l=1}^{K} N_{l,t-1} \sum_{k=1}^{K} \hat{\pi}_{l,k,t} \log \left[\Phi\left(\frac{c_k - b_l f_t - a_l}{\sigma_l}\right) \right. \\
&\qquad \left. - \Phi\left(\frac{c_{k-1} - b_l f_t - a_l}{\sigma_l}\right)\right].
\end{aligned}$$

It follows that the counts $N_{l,t-1}$ and the empirical transition frequencies $\hat{\pi}_{l,k,t}$

for $l, k = 1, \ldots, K$ and $t = 1, \ldots, T$, are summary statistics for the stochastic migration model (conditionally on the initial observations and factor values). These empirical counts and transition frequencies are freely available from websites of either rating agencies, or central banks [see e.g. Gupton, Finger, Bhatia (1997)].

From the cross-sectional micro-density, we get the fixed effects estimators of the factor values given the micro-parameters:

$$\hat{f}_{n,t}(\beta) = \arg\max_{f_t} \sum_{l=1}^{K} N_{l,t-1} \sum_{k=1}^{K} \hat{\pi}_{lk,t} \log \left[\Phi\left(\frac{c_k - b_l f_t - a_l}{\sigma_l} \right) \right.$$
$$\left. - \Phi\left(\frac{c_{k-1} - b_l f_t - a_l}{\sigma_l} \right) \right],$$

for $t = 1, \cdots, T$, and the estimator of the micro-parameters:

$$\hat{\beta}_{nT} = \arg\max_{\beta} \sum_{t=1}^{T} \sum_{l=1}^{K} N_{l,t-1} \sum_{k=1}^{K} \hat{\pi}_{lk,t} \log \left[\Phi\left(\frac{c_k - b_l \hat{f}_{n,t}(\beta) - a_l}{\sigma_l} \right) \right.$$
$$\left. - \Phi\left(\frac{c_{k-1} - b_l \hat{f}_{n,t}(\beta) - a_l}{\sigma_l} \right) \right].$$

The numerical computation of the estimate $\hat{\beta}_{nT}$ involves two nested optimization problems. For given β, the factor approximation $\hat{f}_{n,t}(\beta)$ can be computed by grid search, and then estimate $\hat{\beta}_{nT}$ is computed by applying the Newton-Raphson algorithm. Estimator $\hat{\beta}_{nT}$ is used to get the cross-sectional approximations of the factor values:

$$\hat{f}_{nT,t} = \hat{f}_{n,t}(\hat{\beta}_{nT}). \tag{4.36}$$

Finally, the estimators of the macro-parameters μ, A and Ω are obtained by replacing the factor proxies in the macro-dynamics and applying Maximum Likelihood (ML) on the autoregressive model:

$$\hat{f}_{nT,t} = \mu + A\hat{f}_{nT,t-1} + \varepsilon_t, \quad \varepsilon_t \sim IIN(0, \Omega), \quad t = 1, \cdots, T. \tag{4.37}$$

For this autoregressive model, the ML estimator coincides with the Ordinary Least Squares (OLS) estimator. When the factor is multivariate and the Vector Autoregressive (VAR) model (4.37) involves several equations, the OLS estimators of the components of μ and A are computed equation-by-equation [see e.g. Gourieroux, Jasiak (2001)]. The estimator of Ω is obtained from the sample variance-covariance matrix of the estimated regression residuals $\hat{\varepsilon}_t$. The convergence rates of estimators $\hat{\beta}_{nT}$ and $\hat{\theta}_{nT} = (\hat{\mu}_{nT}, \hat{A}_{nT}, \hat{\Omega}_{nT})$ are $1/\sqrt{nT}$ and $1/\sqrt{T}$, respectively, and their asymptotic variance-covariance matrices are deduced from Proposition 4.1.

v) Approximate linear state space model

An alternative estimation methodology can be introduced by writing the stochastic migration model as an approximate linear state space model, and applying a procedure similar to the one described in Section 4.1. The basic idea is that the qualitative model (4.34) can be "linearized" by considering the canonical factors for each row of the transition matrix as in the ordered probit model in Example 3.4 of Section 3.3. More precisely, let us introduce the cumulated transition probabilities:

$$\begin{aligned} \pi^*_{l,k}(f_t;\beta) &= P[Y_{i,t} \leq k | Y_{i,t-1} = l, f_t; \beta] \\ &= \sum_{h=1}^{k} \pi_{l,h}(f_t;\beta) \\ &= \Phi\left(\frac{c_k - a_l - b_l f_t}{\sigma_l}\right), \end{aligned} \tag{4.38}$$

for $l = 1, \cdots, K$ and $k = 1, \cdots, K-1$. By applying the quantile function of the standard normal distribution to both sides of the above equation, we get:

$$\Phi^{-1}[\pi^*_{l,k}(f_t;\beta)] = \frac{c_k - a_l - b_l f_t}{\sigma_l}. \tag{4.39}$$

These nonlinear transformations of the cumulated transition probabilities play the role of the canonical factors:

$$a_t = vec[a_{l,k,t}], \tag{4.40}$$

where:

$$a_{l,k,t} = \frac{c_k - a_l - b_l f_t}{\sigma_l}, \tag{4.41}$$

is linear w.r.t. f_t. The canonical factors are cross-sectionally approximated by their sample analogues:

$$\hat{a}_{l,k,t} = \Phi^{-1}\left(\sum_{h=1}^{k} \hat{\pi}_{l,h,t}\right). \tag{4.42}$$

For large cross-sectional size n, the estimated factors are such that $\hat{a}_t \overset{d}{\sim} N(a_t, \Sigma_{n,t})$ asymptotically conditional on the factors, and the expression of the estimated asymptotic variance-covariance matrix $\hat{\Sigma}_{n,t}$ is derived in Appendix 4.6. In particular, the block of matrix $\hat{\Sigma}_{n,t}$ corresponding to the canonical factors for row l involves the number $N_{l,t-1}$ of companies in class l at date $t-1$. Then, the parameters can be estimated by applying the Kalman filter on the VGA linear state space model:

State equation:

$$F_t = \mu + AF_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim IIN(0, \Omega);$$

VGA measurement equations:

$$\hat{a}_{l,k,t} = \frac{c_k - a_l}{\sigma_l} - \frac{b_l}{\sigma_l} f_t + u_{l,k,t}, \quad vec(u_{l,k,t}) \sim IIN(0, \hat{\Sigma}_{n,t}),$$

for $l = 1, \cdots, K-1$ and $k = 1, \cdots, K$.

Thus, the nonlinear measurement equations for the individual ratings are approximated by linear measurement equations for suitable cross-sectional aggregates. Contrary to the linear state space models in Section 4.1, the variance of the errors in the measurement equations tend to zero as n tends to infinity. This explains the different rates of convergence for the micro-parameters and the macro-parameters, that are $1/\sqrt{nT}$ and $1/\sqrt{T}$, respectively.

4.5 Application to S&P Migration Data

In this section we present an application of stochastic transition models to rating migration data of S&P [see Gagliardini, Gouriéroux (2005b) for a similar analysis with data of the French central bank].

i) Description of the data

The data consist of $T = 20$ 1-year empirical migration matrices for US firms in the period from 1990 to 2009. The migration matrices are provided by S&P in public reports and are computed on an annual basis from a pool of large and medium size US firms. S&P relies on a rating system based on 8 classes (in the simplest version), denoted AAA, AA, A, BBB, BB, B, C, D. Category AAA corresponds to the lowest risk (i.e. the best credit quality), and category C to the highest risk. Category D corresponds to default. The rating is assigned by expertise, accounting for available information on the firm's business and financial ratios. The pool of monitored firms is constantly updated in time to replace defaulting firms and including new firms. For instance, the distributions of the firms in the pool across rating classes in 1990, and in 2009, are given in Table 4.1.

[Insert Table 4.1: Distribution of firms in the S&P pool across rating classes in 1990 and in 2009.]

The total size n of the pool has almost tripled between 1990 and 2009, passing from about 2000 to almost 6000 firms. The relative importance of rating classes BBB, BB, B, C has increased, while that of classes AAA and AA has decreased. Possible explanations for this phenomenon are either a deterioration of the average credit quality of the firms considered in the pool, or an increased severity in the judgment of the rating agency, or a selectivity when updating the pool.

The transition matrix in 2009 is displayed in Table 4.2. The rows and columns are arranged in order of increasing risk.

[Insert Table 4.2: 1-year transition matrix in 2009]

The matrix in Table 4.2 contains an additional column for firms that are not rated (NR) at the end of year 2009. The firms in category NR likely failed to report their balance sheets, or the reported data had missing information. The percentages of not rated firms are between 3% and 18%, and tend to increase as the quality of the rating at the beginning of 2009 deteriorates. Missing data is mainly due to lack of information disclosure, that may be voluntary or not. Similarly to other rating agencies, S&P does not provide the row corresponding to the transition frequencies from the NR category to the other categories. Hence, the transition matrix in Table 4.2 has to be transformed into a square matrix by imputing the companies in the NR category to the other rating classes. It is usually proceeded by a proportional assignment, that is, for each row the NR companies are assigned to the other rating classes proportionally to the transition frequencies of the latter. It is important to check that the proportional assignment does not induce a selectivity bias. By using migration data of the French central bank, Foulcher et al. (2004) provide evidence that incomplete reporting of balance sheets data is not an indicator of imminent default. This finding supports the practice of proportional assignment and suggests that the increase of the NR percentage in the worst rating classes may be due to the fact that disclosing information is not a priority for firms in a difficult situation.

After the transformation to eliminate the NR category, the transition matrix in 2009 is displayed in Table 4.3.

[Insert Table 4.3: Adjusted 1-year transition matrix in 2009]

The largest transition probabilities appear on the main diagonal of the matrix, pointing to a tendency to stability of the ratings. Moreover, the other transition probabilities that are significantly different from zero correspond in general to transitions involving one bucket. Thus, rating down- or upgrades of more than one rating class over one year are unlikely, unless for the riskiest categories. The last column of the matrix displays the default probabilities. As expected, the default probabilities increase as the rating quality decreases. Finally, the elements in the last row of the transition matrix are all zeros, except a 100% in the last column, reflecting the fact that default is an absorbing state.

Let us now discuss the dynamics of the transition matrices. We focus on up- and down-grade probabilities $u_{l,t} = \sum_{k=l+1}^{K} \hat{\pi}_{l,k,t}$ and $d_{l,t} = \sum_{k=1}^{l-1} \hat{\pi}_{l,k,t}$, respectively, indexed by the initial rating class l [see Section 3.3 iv) for a description of the time series of default probabilities]. We display the time series of up-grade probabilities $u_{l,t}$ in Figure 4.1, and the series of down-grade probabilities $d_{l,t}$ in Figure 4.2. The shaded periods in these figures correspond to recessions in US as identified by the National Bureau of Economic Research (NBER).

[Insert Figure 4.1: Up-grade probabilities]

[Insert Figure 4.2: Down-grade probabilities]

The transition probabilities vary cyclically over time, with similar patterns across rating classes. Peaks of downgrade probabilities, and troughs of up-grade probabilities, are associated with the economic recessions in US. The variations in the down-grade and up-grade probabilities over the time span 1990-2009 have been of the order of $5 - 10\%$ for most rating classes, and of the order of 30% for the riskiest rating class C. The evidence in Figures 4.1 and 4.2 supports the idea that transition probabilities are driven by stochastic factors that are common across rating classes, which is at the core of the stochastic transition model.

ii) Estimation results

We estimate the ordered qualitative stochastic transition model introduced in Section 4.4 ii) with $K = 8$ rating classes by applying the methodology presented in Section 4.4 iv). For compatibility with the notation in Section 4.4, we renumber the rating classes of S&P into 1, 2, ..., 8 with $k = 1$ corresponding to default, $k = 2$ corresponding to C, and so on until $k = 8$ for AAA.

Before applying the estimation procedure, we have to determine the number of factors. For this purpose, in a preliminary step we compute the series of estimated canonical factors $\hat{a}_{l,k,t}$ in (4.42) for k, l, and perform their **principal component analysis**, that is, the spectral decomposition of the $T \times T$ matrix YY', where the row t of matrix Y is given by $\hat{a}_{l,k,t} - \bar{a}_{l,k}$, k, l varying, with $\bar{a}_{l,k} = \frac{1}{T}\sum_t a_{l,k,t}$ (see the Review Appendix A.4). The associated eigenvalues are given in decreasing order in the following table:

81.87	13.92	11.88	9.75	5.58	3.05	2.15	1.66	1.44	0.76
0.50	0.27	0.22	0.10	0.05	0.00	0.00	0.00	0.00	0.00

The first eigenvalue is much larger than the other ones. The second, third and fourth eigenvalues are about of the same magnitude. The components of the standardized eigenvectors (zero sample mean and unitary variance) corresponding to the four largest eigenvalues are displayed in Figure 4.3 as functions of date t.

[Insert Figure 4.3: Eigenvectors from Principal Component Analysis]

The pattern of the eigenvector associated with the largest eigenvalue is compatible with the time evolution of the transition probabilities in Figures 4.1 and 4.2. Indeed, small factor values correspond to large default and downgrade risk (assuming positive factor sensitivities). The troughs in the factor pattern in the periods 1990-91, 2001-2002 and 2008-09 are associated with the troughs in upgrade probabilities and the peaks in downgrade probabilities in Figures 4.1 and 4.2. The pattern of the eigenvector associated with the second largest eigenvalue features a downward trend in the period 1990-1996 and an upward trend in the period 1996-2009, corresponding to an increase, resp. a decrease, in downward risk. Finally, the patterns of the eigenvectors associated with the third and fourth eigenvalues are rather erratic.

Based on the above evidence, we consider a specification with a single factor [see Gagliardini, Gouriéroux (2005b) for a multifactor analysis using

French data]. We impose the identification restrictions $a_5 = 0$, $b_5 = \sigma_5 = 1$ and $c_4 = 0$. These identification restrictions concern the parameters of rating class BBB and the threshold between rating classes BBB and BB [see Section 4.4 iii)]. The estimates of the parameters are displayed in Table 4.4.

[Insert Table 4.4: Parameter estimates]

The upper panel displays the estimates of the threshold parameters c_k. As expected, the estimated thresholds are increasing w.r.t. the rating class index. The middle panel displays the estimates for the parameters in rows $l = 2, 3, \cdots, 8$ of the transition matrix, which correspond to rating classes C, B, ..., AAA in the S&P rating system. The intercepts a_l are increasing with respect to the rating index, which confirms that the underlying quantitative score for credit quality is larger for the less risky rating classes. The parameters b_l are the sensitivities of the different rating classes to the factor. The estimated factor sensitivities are all positive, that is, an increase in the factor improves the underlying quantitative score for credit quality in all rating classes. The volatility parameters σ_l are generally smaller for the riskier rating categories. Finally, the lower panel in Table 4.4 displays the estimates for the parameters of the factor dynamics. The autoregressive coefficient is positive and corresponds to a quite strong persistence of the factor.

Let us now discuss the approximated factor path and the link with the business cycle literature [see also Nickell, Perraudin, Varotto (2000)]. The approximated factor values $\hat{f}_{nT,t}$ in (4.36) are displayed in Figure 4.4. The factor estimates are standardized to get zero mean and unit variance in the sample.

[Insert Figure 4.4: Approximated factor values]

The approximated path of the factor in Figure 4.4 is very close to the components of the eigenvector associated with the largest eigenvalue in the PCA of the estimated canonical factors in Figure 4.3. The cyclical pattern of the factor in Figure 4.4 is clearly related to the business cycle. Indeed, the troughs in the factor pattern are associated with the periods of economic recession in the US. Hence, as expected we find a link between the credit cycle and the business cycle. We notice however some lead-lag effects between the two cycles. For instance, we observe a rather long period of decrease in the credit factor from the peak in 1996 until to trough in 2002 associated with the economic recession in 2001. Instead, in the recent economic crisis, there is a decrease in the credit factor at the inception of the recession in 2008 [see also Gagliardini, Gouriéroux (2005b) for a causality analysis between credit and business cycles using French data].

iii) Estimation of asset and migration correlations

The estimated model can be used to get estimates of asset and migration correlations [De Servigny, Renault (2002), Gagliardini, Gourieroux (2005a)]. Let us consider two firms i and j, that are currently in rating classes l and k, respectively. The asset correlation between these two firms, conditional on the current factor value F_t, is defined as: [3]

$$\rho_{a,lk,t} = corr\left[\log(A_{i,t+1}/L_{i,t+1}), \log(A_{j,t+1}/L_{j,t+1}) | Y_{i,t} = l, Y_{j,t} = k, F_t\right]. \tag{4.43}$$

From equations (4.33) and (4.35), we get:

$$\begin{aligned}\rho_{a,lk,t} &= \frac{b_l b_k V[F_{t+1}|F_t]}{\sqrt{b_l^2 V[F_{t+1}|F_t] + \sigma_l^2}\sqrt{b_k^2 V[F_{t+1}|F_t] + \sigma_k^2}} \\ &= \frac{b_l b_k}{\sqrt{b_l^2 + \frac{1-A^2}{\Omega}\sigma_l^2}\sqrt{b_k^2 + \frac{1-A^2}{\Omega}\sigma_k^2}}.\end{aligned}$$

The asset correlation $\rho_{a,lk,t}$ does not depend on the firms names i and j, but only on their current ratings l and k, since the individual risks are exchangeable within a rating class. Moreover, $\rho_{a,lk,t} \equiv \rho_{a,lk}$ is independent of the

[3] The conditioning set in the definition of asset correlation $\rho_{a,lk,t}$ for firms i and j involves neither the ratings of other firms, nor the past values of ratings and factors. Indeed, this additional information is irrelevant due to the conditional independence of the rating histories given the factor path, and the Markov property of the factor.

date t, since the conditional variance of the Gaussian autoregressive factor is constant. The asset correlation involves both micro parameters b_l, b_k, σ_l^2, σ_k^2 and macro-parameters A, Ω.

The asset correlations $\rho_{a,lk}$ for the different rating classes can be arranged in a symmetric matrix, whose row and column indices l, k correspond to the current firms ratings. The matrix of estimated asset correlations is displayed in Table 4.5.

[Insert Table 4.5: Matrix of estimated asset correlations]

Estimated asset correlations range between about 3% for firms within rating class AAA, and about 26% for firms within rating class C. Hence, the risks in class AAA are mostly driven by the idiosyncratic component, while the systematic factor impacts significantly the risks in class C. This effect can also be seen by comparing the estimated ratios b_l/σ_l for classes $l = 2$ and $l = 8$ in Table 4.4. However, estimated asset correlations are not monotone with respect to the riskiness of the rating class. Indeed, the estimated asset correlation is rather large and equal to about 20% for class AA, while it is equal to about 10% for classes A, BBB, and BB. This finding can be due to the heterogeneity in these latter classes, whose size is large especially in the last part of the sample (see Table 4.1).

Let us now consider the migration correlations. The upgrade correlation between the future ratings of two firms, conditional on the current ratings and factor value, is defined as:

$$\rho_{u,lk,t} = corr\left[\mathbb{1}_{Y_{i,t+1}=l+1}, \mathbb{1}_{Y_{j,t+1}=k+1} | Y_{i,t} = l, Y_{j,t} = k, F_t\right],$$

where l and k are the current ratings of the two firms. Hence, upgrade correlations are correlations between the indicators for the events of rating upgrades. We show in Appendix C that upgrade correlations can be rewritten in terms of conditional moments and cross-moments of the stochastic migration probabilities given the current factor value:

$$\rho_{u,lk,t} = \frac{Cov\left[\pi_{l,l+1,t+1}, \pi_{k,k+1,t+1}|F_t\right]}{\sqrt{E[\pi_{l,l+1,t+1}|F_t](1-E[\pi_{l,l+1,t+1}|F_t])}\sqrt{E[\pi_{k,k+1,t+1}|F_t](1-E[\pi_{k,k+1,t+1}|F_t])}}, \tag{4.44}$$

where $\pi_{l,l+1,t+1} = \pi_{l,l+1}(F_{t+1})$ are the stochastic upgrade probabilities. As for asset correlations, the upgrade correlations depend on the current ratings of the two firms, but not on their names. Moreover, upgrade correlations depend on the current factor value F_t through the conditional distribution of F_{t+1} given F_t, and not only on the conditional variance of the factor. Hence, upgrade correlations are stochastic and time varying. Finally, we can define similarly downgrade correlations and we have:

$$\begin{aligned} \rho_{d,lk,t} &= corr\left[\mathbb{1}_{Y_{i,t+1}=l-1}, \mathbb{1}_{Y_{j,t+1}=k-1}|Y_{i,t}=l, Y_{j,t}=k, F_t\right] \\ &= \frac{Cov\left[\pi_{l,l-1,t+1}, \pi_{k,k-1,t+1}|F_t\right]}{\sqrt{E[\pi_{l,l-1,t+1}|F_t](1-E[\pi_{l,l-1,t+1}|F_t])}\sqrt{E[\pi_{k,k-1,t+1}|F_t](1-E[\pi_{k,k-1,t+1}|F_t])}}, \end{aligned} \tag{4.45}$$

where $\pi_{l,l-1,t+1} = \pi_{l,l-1}(F_{t+1})$ are the stochastic downgrade probabilities. Migration correlations involve the micro-parameters through the transition probabilities, and the macro-parameters through the conditional expectations given the current factor value.

We display in Tables 4.6 and 4.7 the matrices of estimated upgrade and downgrade correlations. The current factor value is $F_t = -2.44$, which is the approximated factor value for 2009 found in Section ii).

[Insert Table 4.6: Matrix of estimated upgrade correlations]

[Insert Table 4.7: Matrix of estimated downgrade correlations]

The conditional expectations with respect to the factor are computed by Monte-Carlo integration based on $1,000,000$ repetitions (see Review Appendix A.1). Estimated upgrade correlations are much smaller than asset correlations, and range between 0.04% for rating class AAA and about 2%

for rating class C. Moreover, estimated upgrade correlations are monotonically increasing with respect to the riskiness of the rating class. Estimated downgrade correlations are typically larger than estimated upgrade correlations. For instance, the estimated downgrade correlation is almost 5% for two firms in rating class C. The impact of the systematic factor is asymmetric with respect to downside and upside risk, and is more pronounced for the former.

4.6 Summary

In a general dynamic framework, we have to account for both micro- and macro-dynamics. We have developed efficient estimation methods for estimating micro- and macro-parameters and shown that they have different rates of convergence. We have also explained how to reconstitute the unobservable dynamic factors. The approach has been applied to the dynamic analysis of corporate rating by means of a stochastic migration model, as recommended by Basel 2 regulation.

REFERENCES

Arellano, M., and S., Bonhomme (2009): "Robust Priors in Nonlinear Panel Data Models", Econometrica, 77, 489-536.

Basel Committee on Banking Supervision (2001): "The New Basel Capital Accord", Consultative Document of the Bank for International Settlements, April 2001, Part 2: Pillar 1.

Basel Committee on Banking Supervision (2003): "The New Basel Capital Accord", Consultative Document of the Bank for International Settlements, April 2003, Part 3: The Second Pillar.

Bartholomew, D. (1982): *Stochastic Models for Social Processes*, 3rd edition, Wiley.

De Servigni, A., and O., Renault (2002): " Default Correlation: Empirical Evidence", Standard & Poor's DP.

Feng, D., Gouriéroux, C., and J., Jasiak (2008): "The Ordered Qualitative Model for Credit Rating Transitions", Journal of Empirical Finance, 15, 111-130.

Foulcher, S., Gouriéroux, C., and A., Tiomo (2004): "The Term Structure of Defaults and Ratings", Insurance and Risk Management, 72, 207-276.

Gagliardini, P., and C., Gouriéroux (2005a): "Migration Correlation: Definition and Efficient Estimation", Journal of Banking and Finance, 29, 865-894.

Gagliardini, P., and C., Gouriéroux (2005b): "Stochastic Migration Models with Application to Corporate Risk", Journal of Financial Econometrics, 3, 188-226.

Gagliardini, P., and C., Gouriéroux (2009): "Efficiency in Large Dynamic Panel Models with Common Factor", CREST DP.

Gordy, M., and E., Heitfield (2002): "Estimating Default Correlation from Short Panels of Credit Rating Performance Data", Working Paper, Federal Reserve Board.

Gouriéroux, C., and J., Jasiak (2001): *Financial Econometrics. Problems, Models and Methods*, Princeton University Press.

Gouriéroux, C., and A., Monfort (1995): *Statistics and Econometric Models*, Cambridge University Press, Cambridge.

Gupton, G., Finger, C., and M., Bhatia (1997): "CreditMetrics", Technical Report, The Riskmetrics Group.

Jensen, J. (1995): *Saddlepoint Approximations*, Clarendon Press, Oxford.

Lancaster, T. (2000): "The Incidental Parameter Problem Since 1948", Journal of Econometrics, 95, 391-413.

Neyman, J., and E., Scott (1948): "Consistent Estimates Based on Partially Consistent Observations", Econometrica, 16, 1-31.

Nickell, P., Perraudin, W., and S., Varotto (2000): "Stability of Rating Transitions", Journal of Banking and Finance, 24, 203-227.

4.7 Appendix A: Asymptotic Variance-Covariance Matrix of the Transition Frequencies

Let us denote $\hat{\pi}_{l,t} = (\hat{\pi}_{l,1,t}, \ldots, \hat{\pi}_{l,K,t})'$, $l = 1, \ldots, K$, $t = 1, \ldots, T$ the rows of the empirical transition matrices. These rows are asymptotically independent conditionally on the factor history such that [see e.g. Bartholomew (1982)]:

$$\sqrt{n}(\hat{\pi}_{l,t} - \pi_{l,t}) \xrightarrow{d} N[0,\ \text{diag}\ (\pi_{l,t}) - \pi_{l,t}\pi'_{l,t}],$$

as $n \to \infty$, where n denotes the number of individuals in class l at date $t-1$. The cumulated transition frequencies $\hat{\pi}^*_{l,t} = (\hat{\pi}^*_{l,1,t}, \ldots, \hat{\pi}^*_{l,K-1,t})'$ are also conditionally independent for different rows and dates, with asymptotic Gaussian distribution:

$$\sqrt{n}(\hat{\pi}^*_{l,t} - \pi^*_{l,t}) \xrightarrow{d} N[0, Q(\text{diag}(\pi_{l,t}) - \pi_{l,t}\pi'_{l,t})Q'],$$

where $Q = [q_{l,k}]$ is the $(K-1, K)$ matrix with $q_{l,k} = 1$, if $k \leq l$, and $= 0$, otherwise.

Finally, the rows of the estimated canonical factor:

$$\hat{a}_{l,t} = [\Phi^{-1}(\hat{\pi}^*_{l,1,t}), \ldots, \Phi^{-1}(\hat{\pi}^*_{l,K-1,t})]', \quad l = 1, \ldots, K,\ t = 1, \ldots, T,$$

are also asymptotically independent with asymptotic distribution:

$$\sqrt{n}(\hat{a}_{l,t} - a_{l,t}) \xrightarrow{d} N\left(0, \Delta_{l,t}Q[\text{diag}(\pi_{l,t}) - \pi_{l,t}\pi'_{l,t}]Q'\Delta_{l,t}\right),$$

where $\Delta_{l,t} = \text{diag}\{(\varphi[\Phi^{-1}(\pi_{k,l,t})])^{-1}, k = 1, \cdots, K-1\}$ by applying the δ-method. The estimated asymptotic variance $\hat{\Sigma}_{n,t}$ of $\hat{a}_t$ is such that the block corresponding to row l is:

$$\hat{\Sigma}_{n,l,t} = \frac{1}{N_{l,t-1}}\hat{\Delta}_{l,t}Q[\text{diag}(\hat{\pi}_{l,t}) - \hat{\pi}_{l,t}\hat{\pi}'_{l,t}]Q'\hat{\Delta}_{l,t},$$

where $\hat{\Delta}_{l,t}$ is defined in terms of the empirical transition frequencies.

4.8 Appendix B: Likelihood Expansion and GAML Estimators

i) Expansion of the log-likelihood

We have :

$$l(\underline{y_T};\beta,\theta) = \int \dots \int \exp\left\{ \sum_{t=1}^{T}\sum_{i=1}^{n} \log h(y_{i,t}|y_{i,t-1}, f_t;\beta) + \sum_{t=1}^{T} \log g(f_t|f_{t-1};\theta)\right\} \prod_{t=1}^{T} df_t.$$

Let us now expand the integrand w.r.t. f_t around $\hat{f}_{nt}(\beta)$, $t = 1,\dots,T$, and define:

$$\begin{aligned}
\psi_{nt}(f_t, f_{t-1}) \quad = \quad & \sum_{i=1}^{n} \log h(y_{i,t}|y_{i,t-1}, f_t;\beta) - \sum_{i=1}^{n} \log h(y_{i,t}|y_{i,t-1}, \hat{f}_{nt}(\beta);\beta) \\
& + \frac{1}{2}\sqrt{n}(f_t - \hat{f}_{nt}(\beta))' I_{nt}(\beta)\sqrt{n}(f_t - \hat{f}_{nt}(\beta)) \\
& + \log g(f_t|f_{t-1};\theta) - \log g(\hat{f}_{nt}(\beta)|\hat{f}_{n,t-1}(\beta);\theta).
\end{aligned}$$

Then:

$$\begin{aligned}
l(\underline{y_T};\beta,\theta) \quad = \quad & \prod_{t=1}^{T}\prod_{i=1}^{n} h(y_{i,t}|y_{i,t-1}, \hat{f}_{nt}(\beta);\beta) \prod_{t=1}^{T} g(\hat{f}_{nt}(\beta)|\hat{f}_{n,t-1}(\beta);\theta) \\
& \int \dots \int \exp\left\{ -\frac{1}{2}\sum_{t=1}^{T} \sqrt{n}(f_t - \hat{f}_{nt}(\beta))' I_{nt}(\beta)\sqrt{n}(f_t - \hat{f}_{nt}(\beta))\right\} \\
& \exp\left\{ \sum_{t=1}^{T} \psi_{n,t}(f_t, f_{t-1})\right\} \prod_{t=1}^{T} df_t.
\end{aligned}$$

Let us introduce the change of variable:

$$Z_t = \sqrt{n}[I_{nt}(\beta)]^{1/2}(f_t - \hat{f}_{nt}(\beta)) \Longleftrightarrow f_t = \hat{f}_{nt}(\beta) + \frac{1}{\sqrt{n}}[I_{nt}(\beta)]^{-1/2} Z_t.$$

Then:

$$
\begin{aligned}
& l(\underline{y_T};\beta,\theta) \\
= & \left(\frac{2\pi}{n}\right)^{TL/2} \prod_{t=1}^{T} [\det I_{nt}(\beta)]^{-1/2} \prod_{t=1}^{T}\prod_{i=1}^{n} h(y_{i,t}|y_{i,t-1},\hat{f}_{nt}(\beta);\beta) \prod_{t=1}^{T} g(\hat{f}_{nt}(\beta)|\hat{f}_{n,t-1}(\beta);\theta) \\
& \frac{1}{(2\pi)^{TL/2}} \int \dots \int \exp\left\{-\frac{1}{2}\sum_{t=1}^{T} Z_t' Z_t\right\}
\end{aligned}
$$

$$
\exp\left\{\sum_{t=1}^{T} \psi_{n,t}\left(\hat{f}_{n,t}(\beta)+\frac{1}{\sqrt{n}}[I_{n,t}(\beta)]^{-1/2}Z_t, \hat{f}_{n,t-1}(\beta)+\frac{1}{\sqrt{n}}[I_{n,t-1}(\beta)]^{-1/2}Z_{t-1}\right)\right\} \prod_{t=1}^{T} dZ_t,
$$

where L is the dimension of the factor f_t. Thus, we can write:

$$
\begin{aligned}
l(\underline{y_T};\beta,\theta) = & \left(\frac{2\pi}{n}\right)^{TL/2} \prod_{t=1}^{T} [\det I_{n,t}(\beta)]^{-1/2} \\
& \cdot \prod_{t=1}^{T}\prod_{i=1}^{n} h(y_{i,t}|y_{i,t-1},\hat{f}_{n,t}(\beta);\beta) \prod_{t=1}^{T} g(\hat{f}_{n,t}(\beta)|\hat{f}_{n,t-1}(\beta);\theta) J_{n,T},
\end{aligned}
$$

where:

$$
\begin{aligned}
J_{n,T} = E\Bigg[\exp\Bigg\{\sum_{t=1}^{T} \psi_{n,t}\Bigg(& \hat{f}_{n,t}(\beta)+\frac{1}{\sqrt{n}}[I_{n,t}(\beta)]^{-1/2}Z_t, \\
& \hat{f}_{n,t-1}(\beta)+\frac{1}{\sqrt{n}}[I_{n,t-1}(\beta)]^{-1/2}Z_{t-1}\Bigg)\Bigg\}\Bigg],
\end{aligned}
$$

is an expectation with respect to independent standard normal variables Z_t, $t = 1, \dots, T$. The result in Proposition 4.2 is deduced by expanding up to order $1/n$ the function within the expectation, and taking into account that the odd power moments of a standard normal variable Z are zero, while $E[Z^2] = 1$, $E[Z^4] = 3$ and $E[Z^6] = 15$ [see Gagliardini, Gouriéroux, (2009) for the detailed derivation].

ii) Newton-Raphson expansion of the GA estimator

By definition of the CSA and GA maximum likelihood estimators, we get:

$$\frac{\partial \mathcal{L}^{CSA}}{\partial(\beta',\theta')'}(\hat{\beta}^{CSA},\hat{\theta}^{CSA}) = 0, \quad \frac{\partial \mathcal{L}^{GA}}{\partial(\beta',\theta')'}(\hat{\beta}^{GA},\hat{\theta}^{GA}) = 0.$$

By considering the expansion of the second set of first-order conditions around the CSA estimator, we get:

$$\frac{\partial \mathcal{L}^{GA}}{\partial(\beta',\theta')'}(\hat{\beta}^{CSA},\hat{\theta}^{CSA}) \simeq -\frac{\partial^2 \mathcal{L}^{GA}(\hat{\beta}^{CSA},\hat{\theta}^{CSA})}{\partial(\beta',\theta')'\partial(\beta',\theta')}\left[\begin{pmatrix}\hat{\beta}^{GA}\\ \\ \hat{\theta}^{GA}\end{pmatrix} - \begin{pmatrix}\hat{\beta}^{CSA}\\ \\ \hat{\theta}^{CSA}\end{pmatrix}\right].$$

This is equivalent to:

$$\begin{aligned}
\begin{pmatrix}\hat{\beta}^{GA}\\ \\ \hat{\theta}^{GA}\end{pmatrix} - \begin{pmatrix}\hat{\beta}^{CSA}\\ \\ \hat{\theta}^{CSA}\end{pmatrix} &\simeq \left[-\frac{\partial^2 \mathcal{L}^{GA}(\hat{\beta}^{CSA},\hat{\theta}^{CSA})}{\partial(\beta',\theta')'\partial(\beta',\theta')}\right]^{-1} \frac{\partial \mathcal{L}^{GA}(\hat{\beta}^{CSA},\hat{\theta}^{CSA})}{\partial(\beta',\theta')'} \\
&\simeq \left[-\frac{\partial^2 \mathcal{L}^{CSA}(\hat{\beta}^{CSA},\hat{\theta}^{CSA})}{\partial(\beta',\theta')'\partial(\beta',\theta')}\right]^{-1} \frac{\partial \mathcal{L}^{GA}(\hat{\beta}^{CSA},\hat{\theta}^{CSA})}{\partial(\beta',\theta')'}.
\end{aligned}$$

4.9 Appendix C: Migration Correlations

In this Appendix we prove equations (4.44) and (4.45). By the Law of Iterated Expectation, we have:

$$\begin{aligned}
& E\left[\mathbb{1}_{Y_{i,t+1}=l+1}|Y_{i,t}=l, Y_{j,t}=k, F_t\right] \\
=\;& E\left[E\left[\mathbb{1}_{Y_{i,t+1}=l+1}|Y_{i,t}=l, Y_{j,t}=k, F_{t+1}, F_t\right]|Y_{i,t}=l, Y_{j,t}=k, F_t\right] \\
=\;& E\left[P\left[Y_{i,t+1}=l+1|Y_{i,t}=l, Y_{j,t}=k, F_{t+1}, F_t\right]|Y_{i,t}=l, Y_{j,t}=k, F_t\right] \\
=\;& E\left[\pi_{l,l+1}(F_{t+1})|Y_{i,t}=l, Y_{j,t}=k, F_t\right] = E\left[\pi_{l,l+1}(F_{t+1})|F_t\right].
\end{aligned}$$

By a similar argument, and by using the conditional independence of the rating histories given the factor path, we have:

$$\begin{aligned}
& E\left[\mathbb{1}_{Y_{i,t+1}=l+1}\mathbb{1}_{Y_{j,t+1}=k+1}|Y_{i,t}=l, Y_{j,t}=k, F_t\right] \\
=\;& E\left[P\left[Y_{i,t+1}=l+1, Y_{j,t+1}=k+1|Y_{i,t}=l, Y_{j,t}=k, F_{t+1}, F_t\right]|Y_{i,t}=l, Y_{j,t}=k, F_t\right] \\
=\;& E\left[\pi_{l,l+1}(F_{t+1})\pi_{k,k+1}(F_{t+1})|F_t\right].
\end{aligned}$$

Thus, we get:

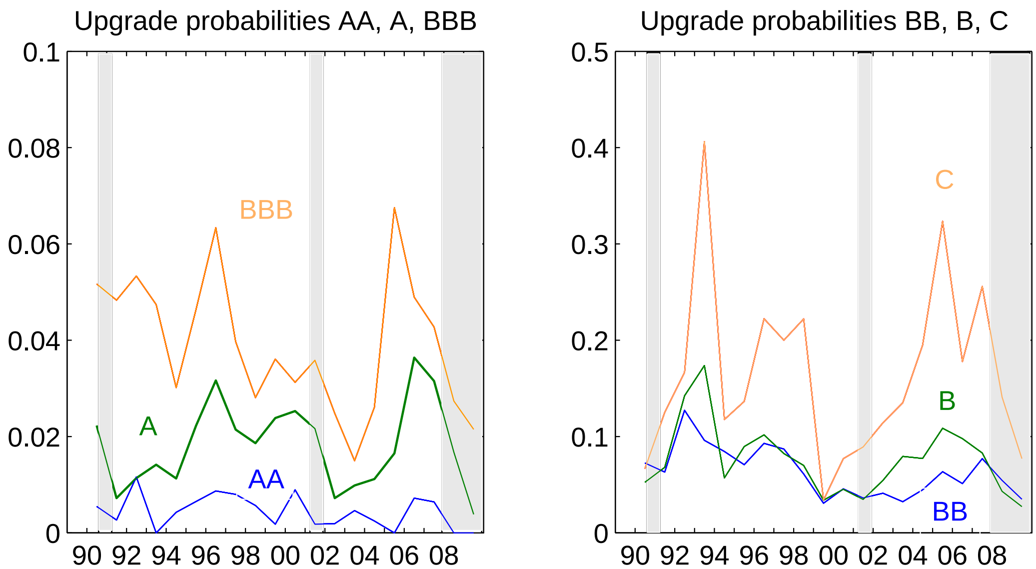
$$V\left[\mathbb{1}_{Y_{i,t+1}=l+1}|Y_{i,t}=l,Y_{j,t}=k,F_t\right]=E\left[\pi_{l,l+1}(F_{t+1})|F_t\right](1-E\left[\pi_{l,l+1}(F_{t+1})|F_t\right]),$$

and:

$$Cov\left[\mathbb{1}_{Y_{i,t+1}=l+1},\mathbb{1}_{Y_{j,t+1}=k+1}|Y_{i,t}=l,Y_{j,t}=k,F_t\right]=Cov\left[\pi_{l,l+1}(F_{t+1}),\pi_{k,k+1}(F_{t+1})|F_t\right].$$

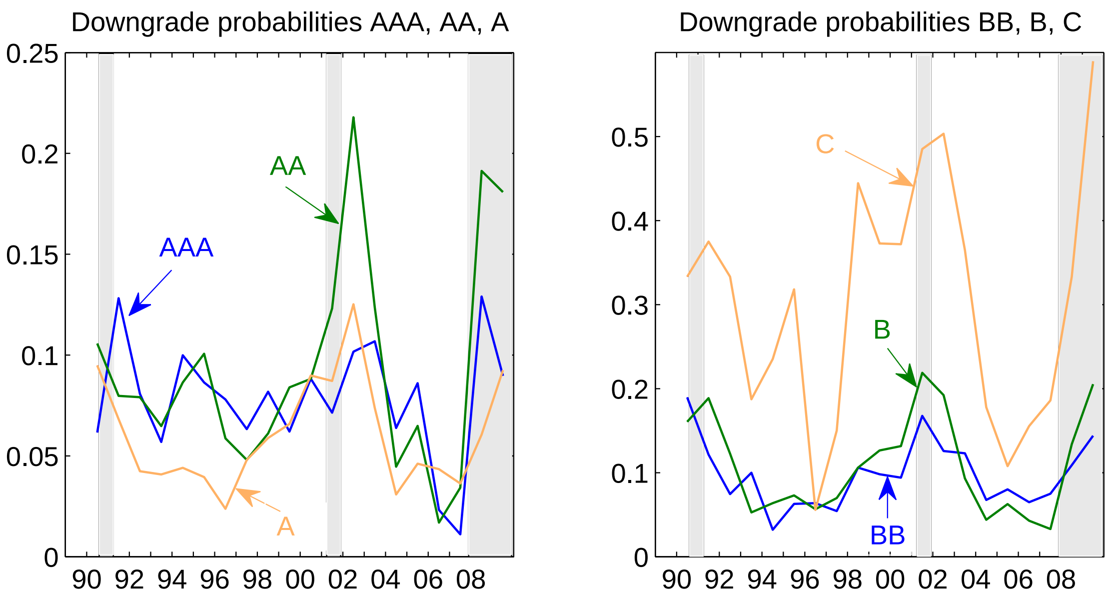
Equation (4.44) follows. The proof of (4.45) is similar.

Figure 4.1: Up-grade probabilities.



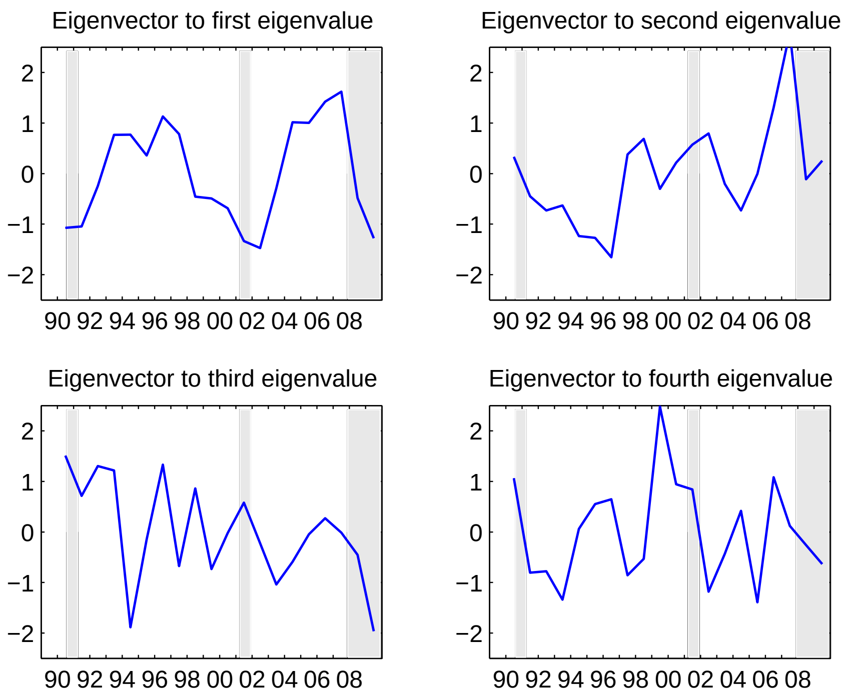
The figure displays the time-series of 1-year up-grade probabilities for rating classes AA, A, BBB (left panel), and BB, B, C (right panel) in the period 1990-2009. Migration probabilities are in percentage. Shaded periods correspond to NBER recessions in US.

Figure 4.2: Down-grade probabilities.



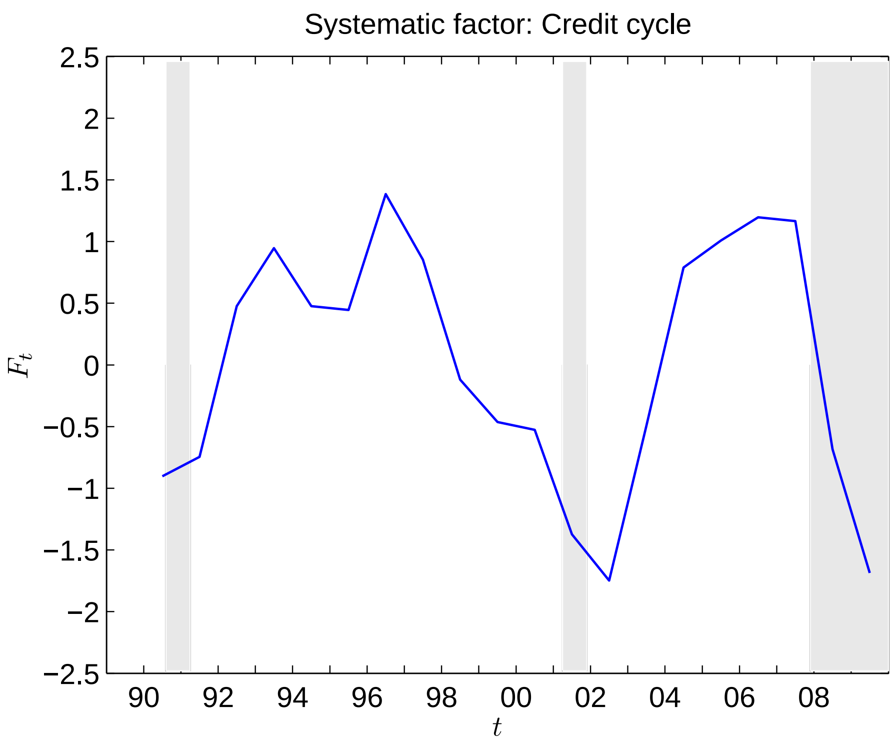
The figure displays the time-series of 1-year down-grade probabilities for rating classes AAA, AA, A (left panel), and BB, B, C (right panel) in the period 1990-2009. Migration probabilities are in percentage. Shaded periods correspond to NBER recessions in US.

Figure 4.3: Eigenvectors from Principal Component Analysis.



The figure displays the patterns of the eigenvectors associated with the four largest eigenvalues in the principal components analysis of the estimated canonical factors. Shaded periods correspond to NBER recessions in US.

Figure 4.4: Approximated factor values.



The figure displays the pattern of the approximated factor values $\hat{f}_{nT,t}$ [see (4.36)] for $t = 1, ..., 20$. The factor estimates are standardized to get zero mean and unit variance in the sample. Shaded periods correspond to NBER recessions in US.

Table 4.1: Distribution of firms in the S&P pool across rating classes in 1990 and in 2009.

	AAA	AA	A	BBB	BB	B	C	Total
1990	147	373	560	347	282	363	48	2120
2009	81	470	1396	1498	1002	1223	190	5860

Table 4.2: 1-year transition matrix.

		2009								
		AAA	AA	A	BBB	BB	B	C	D	NR
	AAA	87.65	8.64	0	0	0	0	0	0	3.71
	AA	0	76.17	15.96	0.64	0.21	0	0	0	7.02
	A	0	0.36	84.67	7.74	0.43	0.29	0	0.21	6.3
2008	BBB	0	0	2	83.71	5.94	0.8	0.20	0.53	6.81
	BB	0	0	0	3.09	72.95	11.48	0.60	0.70	11.18
	B	0	0	0.16	0	2.29	69.34	8.42	10.14	9.65
	C	0	0	0	0	0	6.32	27.37	48.42	17.89
	D	0	0	0	0	0	0	0	100	0

1-year transition matrix for 2009. Transition probabilities are in percentage. Rating classes are ordered from AAA (lowest risk) to D (default). The column NR corresponds to firms that are not rated at the end of 2009.

Table 4.3: Adjusted 1-year transition matrix.

		2009							
		AAA	AA	A	BBB	BB	B	C	D
	AAA	91.02	8.98	0	0	0	0	0	0
	AA	0	81.92	17.16	0.69	0.23	0	0	0
	A	0	0.38	90.36	8.26	0.46	0.31	0	0.22
2008	BBB	0	0	2.15	89.83	6.37	0.86	0.21	0.57
	BB	0	0	0	3.48	82.13	12.93	0.68	0.79
	B	0	0	0.18	0	2.53	76.75	9.32	11.22
	C	0	0	0	0	0	7.70	33.33	58.97
	D	0	0	0	0	0	0	0	100

Adjusted 1-year transition matrix for 2009. Transition probabilities are in percentage. Rating classes are ordered from AAA (lowest risk) to C (default).

Table 4.4: Parameter estimates

$c_1 = -2.635$	$c_2 = -2.517$	$c_3 = -1.818$	$c_4 = 0$	$c_5 = 6.304$
$c_6 = 22.679$	$c_7 = 64.906$			

$a_2 = -2.947$	$b_2 = 0.084$	$\sigma_2 = 0.045$
$a_3 = -2.888$	$b_3 = 0.230$	$\sigma_3 = 0.141$
$a_4 = -2.010$	$b_4 = 0.333$	$\sigma_4 = 0.355$
$a_5 = 0$	$b_5 = 1$	$\sigma_5 = 1$
$a_6 = 4.524$	$b_6 = 2.849$	$\sigma_6 = 2.397$
$a_7 = 9.849$	$b_7 = 8.586$	$\sigma_7 = 5.358$
$a_8 = 78.067$	$b_8 = 8.063$	$\sigma_8 = 15.084$

$\mu = 1.108$	$A = 0.628$	$\Omega = 0.062$

Estimated parameters for the factor ordered probit model. Thresholds c, intercepts a, factor sensitivities b, volatilities σ and parameters of the factor dynamics μ, A, Ω are displayed.

Table 4.5: Matrix of estimated asset correlations.

	AAA	AA	A	BBB	BB	B	C
AAA	2.83	7.66	5.96	5.11	4.81	7.76	8.69
AA	7.66	20.73	16.15	13.84	13.04	21.01	23.54
A	5.96	16.15	12.58	10.78	10.15	16.36	18.33
BBB	5.11	13.84	10.78	9.24	8.70	14.02	15.71
BB	4.81	13.04	10.15	8.70	8.20	13.21	14.80
B	7.76	21.01	16.36	14.02	13.21	21.28	23.84
C	8.69	23.54	18.33	15.71	14.80	23.84	26.71

This Table displays the estimated asset correlations $\rho_{a,lk}$ between two firms (in percentage). Row and column indices l and k, respectively, correspond to the current rating classes of the firms.

Table 4.6: Matrix of estimated upgrade correlations.

	AA	A	BBB	BB	B	C
AA	0.04	0.07	0.08	0.10	0.17	0.28
A	0.07	0.11	0.13	0.15	0.26	0.45
BBB	0.08	0.13	0.16	0.19	0.32	0.54
BB	0.10	0.15	0.19	0.22	0.37	0.63
B	0.17	0.26	0.32	0.37	0.64	1.09
C	0.28	0.45	0.54	0.63	1.09	1.85

This Table displays the estimated upgrade correlations $\rho_{u,lk,t}$ between two firms (in percentage). Row and column indices l and k, respectively, correspond to the current rating classes of the firms. The current factor value is equal to the approximated factor value for year 2009.

Table 4.7: Matrix of estimated downgrade correlations.

	AAA	AA	A	BBB	BB	B	C
AAA	0.14	0.49	0.29	0.21	0.25	0.34	0.82
AA	0.49	1.79	1.06	0.75	0.92	1.21	2.93
A	0.29	1.06	0.63	0.45	0.55	0.72	1.74
BBB	0.21	0.75	0.45	0.32	0.39	0.51	1.24
BB	0.25	0.92	0.55	0.39	0.48	0.63	1.53
B	0.34	1.21	0.72	0.51	0.63	0.83	2.02
C	0.82	2.93	1.74	1.24	1.53	2.02	4.95

This Table displays the estimated downgrade correlations $\rho_{d,lk,t}$ between two firms (in percentage). Row and column indices l and k, respectively, correspond to the current rating classes of the firms. The current factor value is equal to the approximated factor value for year 2009.

Chapter 5

Prediction, Filtering and Basket Derivative Pricing

We consider in this chapter an exchangeable set of individual histories, with only macro-dynamics. From Chapter 1, the dynamics is specified by means of a state space model. The measurement equations are defined by the conditional pdf $h(y_{it}|f_t)$ of the individual variables given the common factor. The transition equation is defined by the conditional pdf $g(f_t|f_{t-1})$ of the current factor value given its own past. For expository purpose, we focus in this chapter on a single factor model, but the results can be generalized to multiple-factor models.

As usual, the joint distribution of individual histories involves multiple integrals. Such multiple integrals are also involved when predicting future values of the individual variables, or when trying to reconstitute the unobserved factor values from the observed individual variables, the so-called filtering problem (see Review A.5). Granularity approximations for prediction and filtering problems are the subject of this chapter.

The first section considers granularity adjustments for factor filtering and extends the example considered in Section 2.2 to a general framework. Then, the result is used to deduce the granularity adjustment when we are interested in the prediction of a function of future values of the individual variables.

The results are illustrated in Section 5.2 by various examples. Under the assumption of absence of arbitrage opportunities in the market, the problem of derivative pricing is a prediction problem after an appropriate discounting (see Review B.2 on arbitrage). This explains why the results of Section 5.1 can be used to derive approximate prices for derivatives written on an homogenous basket of individual risks. We give in Section 5.3 examples of such derivatives recently introduced on financial markets. They include Basket Default Swap (BDS), derivatives written on the iTraxx index, longevity bonds, or Mortality Linked Securities (MLS). The corresponding approximated pricing formulas are given and discussed in Section 5.4. Section 5.5 explains how to introduce appropriately designed derivatives for hedging a common risk. Finally, in Section 5.6 we present a numerical illustration for the approximate pricing of BDS.

5.1 Approximate Prediction Formulas

We first derive an approximation at order $1/n$ of the predictive distribution of f_t given all individual histories up to date t. Then, this formula is used to derive the prediction of future values of the factor and of the individual variables.

i) Approximate filtering

Let us assume known the individual histories up to date t. They are denoted by $\underline{y_{i,t}}$, for $i = 1, \ldots, n$. The cross-sectional maximum likelihood estimate:

$$\hat{f}_{nt} = \arg\max_{f_t} \sum_{i=1}^{n} \log h(\underline{y_{i,t}}|f_t), \tag{5.1}$$

provides a first approximation of the unknown factor value. This approximation is based on the cross-sectional information, but neglects the information contained in past observations. The proposition below explains how it can be improved. It provides a result valid for the predictive distribution itself, which can be characterized by the knowledge of its **Laplace transform**, that is, by the knowledge of the prediction of any exponential transform of f_t. This transformation can be real (moment generating function), or complex (characteristic function). In the latter case, it provides directly the predictions of sine and cosine transforms of f_t, and then of any function of f_t by Fourier inversion.

The proposition below is derived in Appendix 5.7 i).

Proposition 5.1: *We have*

$$
\begin{aligned}
& E[\exp(uf_t)|\underline{y_{1,t}}, \dots, \underline{y_{n,t}}, \underline{f_{t-1}}] \\
= \; & E[\exp(uf_t)|\underline{y_{1,t}}, \dots, \underline{y_{n,t}}] + o(1/n) \\
= \; & \exp\{u[\hat{f}_{nt} + \frac{1}{n} I_{nt}^{-1} \frac{\partial \log g}{\partial f_t}(\hat{f}_{nt}|\hat{f}_{n,t-1}) + \frac{1}{2n} I_{nt}^{-2} K_{nt}] + \frac{1}{2n} I_{nt}^{-1} u^2 + o(1/n)\},
\end{aligned}
$$

where $I_{n,t} = -\frac{1}{n}\sum_{i=1}^{n} \frac{\partial^2 \log h}{\partial f_t^2}(y_{i,t}|\hat{f}_{nt})$ *and* $K_{nt} = \frac{1}{n}\sum_{i=1}^{n} \frac{\partial^3 \log h}{\partial f_t^3}(y_{it}|\hat{f}_{n,t})$.

The predictive distribution of f_t, that is, the filtering distribution, depends on the individual histories by means of a small number of summary statistics, that are $\hat{f}_{nt}$, $\hat{f}_{n,t-1}$, I_{nt}, which approximate the cross-sectional information matrix, and K_{nt}, which is a component in the bias at order $1/n$ of the cross-sectional maximum likelihood estimator.

We immediately deduce from Proposition 5.1 the following Corollaries:

Corollary 5.2: *The lagged factor values are not informative at order* $1/n$ *to predict* f_t.

This is a direct consequence of the first equality in Proposition 5.1.

Corollary 5.3: *At order* $1/n$*, the filtering distribution is Gaussian:*

$$
N\left(\hat{f}_{nt} + \frac{1}{n}[I_{nt}^{-1}\frac{\partial \log g}{\partial f_t}(\hat{f}_{nt}|\hat{f}_{n,t-1}) + \frac{1}{2} I_{nt}^{-2} K_{nt}], \frac{1}{n} I_{nt}^{-1}\right).
$$

Indeed the Laplace transform of the Gaussian distribution $N(m, \sigma^2)$ is $\exp[um + u^2\sigma^2/2]$ and the result is deduced from Proposition 5.1.

Proposition 5.1 and its Corollaries show that the initial non Gaussian filter can be replaced by an approximate Gaussian filter. This approximate Gaussianity is a numerical result due to a Laplace approximation of the integral underlying the conditional expectation in Proposition 1 [see Appendix 5.7 i)], and not a consequence of a Central Limit Theorem.

When n diverges to infinity, the Gaussian distribution in Corollary 5.3 becomes degenerate, with mean $\hat{f}_{n,t}$ and zero variance. For finite n, the GA has two components, that concern the mean and variance of the approximately Gaussian filtering distribution, respectively. The macrodynamics appears by means of the adjustment of the mean [see the term $\frac{\partial \log g}{\partial f_t}(\hat{f}_{nt}|\hat{f}_{n,t-1})$], whereas cross-sectional effects impact both the mean and variance GA.

From the approximate filtering distribution, we deduce an approximation of the prediction of any smooth function $a(f_t)$ of the factor [see Appendix 5.7 ii)].

Corollary 5.4: *For any twice differentiable function a, we have:*

$$\begin{aligned} E[a(f_t)|\underline{y_{1,t}}, \dots, \underline{y_{n,t}}] &= a(\hat{f}_{nt}) + \frac{1}{n}\frac{da}{df}(\hat{f}_{nt})[I_{nt}^{-1}\frac{\partial \log g}{df_t}(\hat{f}_{nt}|\hat{f}_{n,t-1}) + \frac{1}{2}I_{nt}^{-2}K_{nt}] \\ &\quad + \frac{1}{2n}\frac{d^2 a}{df^2}(\hat{f}_{nt})I_{nt}^{-1} + o(1/n). \end{aligned}$$

The filtering formula above is a kind of Ito's formula [Ito (1951)] for prediction, with both the effect of mean and variance GA at order $1/n$.

ii) Approximate prediction

Let us now consider the prediction of a function $\alpha(y_{1,t+h}, \dots, y_{n,t+h}, f_{t+h})$, say, performed at time t. By the iterated expectation theorem, we know that:

$$E[\alpha(y_{1,t+h}, \dots, y_{n,t+h}, f_{t+h})|\underline{y_{1,t}}, \dots, \underline{y_{n,t}}, \underline{f_{t-1}}] = E[\alpha^*(f_{t+h})|\underline{y_{1,t}}, \dots, \underline{y_{n,t}}, \underline{f_{t-1}}],$$

where $\alpha^*(f_{t+h}) = E[\alpha(y_{1,t+h},\ldots,y_{n,t+h},f_{t+h})|\underline{y_{1,t+h-1}},\ldots,\underline{y_{n,t+h-1}},\underline{f_{t+h}}]$ depends on the conditioning variables by means of f_{t+h} only.

Let us denote:

$$\alpha^*(h,f_t) = E[\alpha^*(f_{t+h})|f_t] = E[\alpha^*(f_{t+h})|\underline{y_{1,t}},\ldots,\underline{y_{n,t}},\underline{f_t}]. \qquad (5.2)$$

By the iterated expectation theorem, we deduce the following result:

Proposition 5.5: *We have:*

$$E[\alpha(y_{1,t+h},\ldots,y_{n,t+h},f_{t+h})|\underline{y_{1,t}},\ldots,\underline{y_{n,t}},\underline{f_{t-1}}] = E[\alpha^*(h,f_t)|\underline{y_{1,t}},\ldots\underline{y_{n,t}},\underline{f_{t-1}}],$$

where $\alpha^*(h,f)$ is given in (5.2).

Thus, we can apply Corollary 5.4 to deduce the GA for any predictor.

Corollary 5.6: *At order $1/n$ the predictor of $\alpha(y_{1,t+h},\ldots,y_{n,t+h},f_{t+h})$ is given by:*

$$\alpha^*(h,\hat{f}_{nt}) + \frac{1}{n}\frac{\partial\alpha^*}{\partial f}(h,\hat{f}_{nt})[I_{nt}^{-1}\frac{\partial\log g}{\partial f_t}(\hat{f}_{nt}|\hat{f}_{n,t-1}) + \frac{1}{2}I_{nt}^{-2}K_{nt}] + \frac{1}{2n}\frac{\partial^2\alpha^*}{\partial f^2}(h,\hat{f}_{nt})I_{nt}^{-1}.$$

The different predictions are simply derived by combining the quantities $\alpha^*(h,\hat{f}_{nt})$, $\frac{\partial\alpha^*}{\partial f}(h,\hat{f}_{nt})$, $\frac{\partial^2\alpha^*(h,\hat{f}_{nt})}{\partial f^2}$ with weights independent of the prediction horizon and of the quantity to be predicted.

iii) Approximate linear state space models

When the factor dynamics is Gaussian autoregressive, the approximate filtering distribution in Corollary 5.3 coincides up to order $o(1/n)$ with the filtering distribution derived from the Kalman filter applied to an approximate linear state space model (see Review A.5 on Kalman filter). Specifically, let us assume that factor (f_t) follows a Gaussian autoregressive process:

$$f_t = \mu + \gamma f_{t-1} + \eta u_t, \quad u_t \sim IIN(0,1), \qquad (5.3)$$

where the autoregressive coefficient γ is such that $|\gamma| < 1$. Then, let us consider the linear state space model that is defined by the measurement equation:

$$\xi_{n,t} = f_t + \frac{1}{\sqrt{n}} I_{n,t}^{-1/2} \varepsilon_t, \quad \varepsilon_t \sim IIN(0,1), \tag{5.4}$$

where $\xi_{n,t} = \hat{f}_{n,t} + \frac{1}{2n} I_{n,t}^{-2} K_{n,t}$, and transition equation (5.3). In the measurement equation, the variable $\xi_{n,t}$ is the cross-sectional factor approximation $\hat{f}_{n,t}$ adjusted by a bias correction term at order $1/n$, while the variance of the error is $\frac{1}{n} I_{n,t}^{-1}$ and vanishes when n diverges to infinity. In Appendix 5.7 iii) we show that the filtering distribution of factor f_t obtained by applying the Kalman filter to the linear state space model (5.3)-(5.4) equals the Gaussian distribution in Corollary 5.3 up to terms of order $o(1/n)$. Moreover, the results in Gagliardini and Gourieroux (2010) show that the approximate linear state space model (5.3)-(5.4) can be used to compute estimators of the macro-parameters μ, γ, η, that are asymptotically equivalent to GA maximum likelihood estimators (see also Chapters 3 and 4 for similar results when variables $y_{i,t}$ are qualitative). Hence, by appropriately linearizing the original nonlinear state space model, we can compute jointly macro-parameter estimates and filtering distributions by applying the standard Kalman filter.

5.2 Examples

In the standard cases, the GA adjustments for the mean and variance have simple expressions. However, function $\alpha^*(h, f)$ can be difficult to derive for large horizon h and complicated function α. We will see in Section 5.5 a case in which it is easily approximated at order $1/n$. We consider below the computations of mean and variance GA coefficients in various examples.

i) Gaussian linear factor model

The individual variables are real valued, such that:

$$y_{i,t} = a + b f_t + \sigma u_{it}, \tag{5.5}$$

where the error terms u_{it} are $IIN(0,1)$ conditional on factor f_t. Since the factor f_t is unobservable, the factor can be transformed such that we have $a = 0$ and $b = 1$. Then, the microdensity is:

$$\prod_{i=1}^{n} h(y_{i,t}|f_t) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\{-\frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_{it} - f_t)^2\}.$$

The cross-sectional maximum likelihood estimator is $\hat{f}_{nt} = \frac{1}{n}\sum_{i=1}^{n} y_{i,t}$, and we have $I_{n,t} = 1/\sigma^2$, $K_{n,t} = 0$.

ii) Stochastic volatility model with factor

The individual observations are such that:

$$y_{i,t} = f_t^{1/2} u_{i,t}, \tag{5.6}$$

where factor (f_t) is a positive Markov process and the error terms $u_{i,t}$ are $IIN(0,1)$ conditional on factor f_t. The microdensity is:

$$\prod_{i=1}^{n} h(y_{i,t}|f_t) = \frac{1}{(2\pi)^{n/2}} \exp\{-\frac{n}{2}\log f_t - \frac{1}{2f_t}\sum_{i=1}^{n} y_{it}^2\}.$$

The cross-sectional maximum likelihood estimator of f_t is $\hat{f}_{nt} = \frac{1}{n}\sum_{i=1}^{n} y_{i,t}^2$. It is equal to a cross-sectional realized variance. Moreover, we have $I_{nt} = 1/(2\hat{f}_{nt}^2)$ and $K_{nt} = 2/(\hat{f}_{nt}^3)$.

iii) Dichotomous qualitative model with factor

The individual variables are dichotomous qualitative; they are independent conditional on the value of a common factor f_t with Bernoulli distribution such that $y_{i,t} \sim \mathcal{B}(1, f_t)$. The factor f_t takes values in the interval $(0,1)$. The cross-sectional estimator of f_t is $\hat{f}_{n,t} = \bar{y}_{n,t} = \frac{1}{n}\sum_{i=1}^{n} y_{i,t}$, and we have:

$$I_{nt} = 1/[\bar{y}_{n,t}(1 - \bar{y}_{n,t})], \quad K_{nt} = 2(1 - 2\bar{y}_{n,t})/[\bar{y}_{n,t}(1 - \bar{y}_{n,t})]^2. \tag{5.7}$$

iv) Gamma model with factor

In this case the individual observations are independent, conditional on factor f_t with common distribution $\gamma(f_t, \lambda)$; thus, factor f_t is a stochastic degree of freedom, with positive real values. The microdensity is:

$$\prod_{i=1}^{n} h(y_{i,t}|f_t) = \frac{1}{\Gamma(f_t)^n} \exp(-\lambda \sum_{i=1}^{n} y_{i,t})(\prod_{i=1}^{n} y_{i,t})^{f_t-1} \lambda^{nf_t} \mathbb{1}_{\min_i y_{i,t}>0},$$

where Γ denotes the gamma function. The cross-sectional maximum likelihood estimator of f_t, derived by maximizing $\prod_{i=1}^{n} h(y_{i,t}|f_t)$ with respect to f_t, does not admit a closed form expression. It is given by:

$$\hat{f}_{n,t} = \psi^{-1}[\frac{1}{n} \sum_{i=1}^{n} \log y_{i,t} + \log \lambda],$$

where $\psi(s) = \dfrac{d \log \Gamma(s)}{ds}$ is the digamma function, and we have:

$$I_{n,t} = \frac{d\psi}{ds}(\hat{f}_{n,t}) \text{ and } K_{n,t} = -\frac{d^2\psi}{ds^2}(\hat{f}_{n,t}).$$

v) Beta model with factor

The individual observations take value in the interval (0,1). They are independent conditionally on factor f_t, with density:

$$h(y_{i,t}|f_t) = \frac{\Gamma(f_t)}{\Gamma(\alpha f_t)\Gamma[(1-\alpha)f_t]} y_{i,t}^{\alpha f_t - 1}(1 - y_{i,t})^{(1-\alpha)f_t - 1} \mathbb{1}_{0<y_{i,t}<1}, \tag{5.8}$$

where α is a scalar parameter in $(0, 1)$ and f_t a positive factor. For this beta distribution, the conditional mean $E(y_{i,t}|f_t) = \alpha$ is constant. Moreover, the conditional variance of a variable on $[0, 1]$ is upper bounded:

$$V(y_{i,t}|f_t) \leq E(y_{i,t}|f_t)[1 - E(y_{i,t}|f_t)] = \alpha(1-\alpha),$$

and the upper bound is reached when the total mass is distributed on the two-points set $\{0, 1\}$. It is easily checked that:

$$f_t + 1 = \alpha(1-\alpha)/V(y_{i,t}|f_t),$$

measures the concentration of the distribution. Thus, we get a beta model with a stochastic concentration parameter.

As in example iv), the cross-sectional maximum likelihood does not admit a simple closed form expression. It is given by:

$$\hat{f}_{n,t} = \psi_\alpha^{-1}\Big(\frac{1}{n}\sum_{i=1}^{n}[\alpha \log y_{i,t} + (1-\alpha)\log(1-y_{i,t})]\Big),$$

where $\psi_\alpha(s) = \alpha\psi(\alpha s) + (1-\alpha)\psi[(1-\alpha)s] - \psi(s)$, and $\psi(s) = \dfrac{d\log\Gamma(s)}{ds}$. Moreover, we have:

$$I_{n,t} = \frac{d\psi_\alpha}{ds}(\hat{f}_{n,t}) \text{ and } K_{n,t} = -\frac{d^2\psi_\alpha}{ds^2}(\hat{f}_{n,t}).$$

5.3 Basket Derivatives

A basket derivative is a derivative written on a large number of individual risks $y_{i,t}$, for $i = 1,\dots,n$. These risks can correspond to individual asset returns, or simply to individual risky events not necessarily traded on financial markets, such as human lifetimes [see the example of longevity in Section 3.4 iii)]. Let us denote by t the current date; a European basket derivative with time-to-maturity h will pay the contractual amount $a(y_{1,t+h},\dots,y_{n,t+h})$, say, at date $t+h$. Its current price is denoted by $\pi_t(a,h)$.

Various basket derivatives have recently been introduced on markets for securitized products, with the aim of making easier an appropriate hedging of some common risks.

i) Basket Default Swap

Let us consider at date t a set of loans $i = 1,\dots,n$ called the basket. A **Basket Default Swap (BDS)** with maturity $t+h$, will pay 1$, say, at time $t+h$, if the proportion of loans with default at date $t+h$ in the basket is larger than a given contractual threshold α, with $\alpha \in (0,1)$. Thus, the design of the BDS is characterized by the composition of the basket, the maturity and the threshold on the default frequency.

Let us represent the individual loan histories by means of the default indicator $y_{i,t}$, such that $y_{i,t} = 1$, if the loan is defaulted at time t, and $y_{i,t} = 0$, otherwise. The payoff of the BDS is:

$$a(y_{i,t+h}, \dots, y_{n,t+h}; \alpha) = \mathbb{1}_{\bar{y}_{n,t+h} > \alpha},$$

where $\bar{y}_{n,t+h} = \dfrac{1}{n} \sum_{i=1}^{n} y_{i,t+h}$.

Let us assume that the basket is homogenous, and that, at a given date, the default indicators of the loans which are still alive are independent, with identical Bernoulli distributions $\mathcal{B}(1, f_t)$, say, conditionally on a common factor f_t. This factor is the stochastic default probability. In the limiting case of a basket of infinite size, we have $\bar{y}_{n,t+1} \sim f_{t+1}$, $\bar{y}_{n,t+2} \sim f_{t+1} + (1 - f_{t+1}) f_{t+2}$, ... Thus, the BDS are derivatives to hedge extreme values of f_{t+1}, $f_{t+1} + (1 - f_{t+1}) f_{t+2}$, ... When n is large, but finite, the interpretation of the BDS as an hedging product is similar, but the insurance against common factor movements cannot be perfect, since the factor values are never observed.

ii) CDO Tranche

A **CDO tranche** is also based on a contractual basket of loans, with payoff at $t + h$ of the type:

$$a(y_{1,t+h}, \dots, y_{n,t+h}; \alpha_1, \alpha_2) = (\bar{y}_{n,t+h} - \alpha_1)^+ - (\bar{y}_{n,t+h} - \alpha_2)^+,$$

where $\alpha_1 < \alpha_2$ are called the **attachment** and **detachment points**, respectively, and $X^+ = \max(X, 0)$. The payoff of the CDO tranche as a function of the default frequency at maturity is displayed in Figure 5.1.

[Insert Figure 5.1: Payoff of a CDO tranche]

The payoff function is nonlinear and corresponds to the payoff of a portfolio which is long in a call option written on $\bar{y}_{n,t+h}$ with strike α_1 and short in a call option with strike α_2.

iii) Derivatives on iTraxx

The **Credit Default Swaps** are life insurance contracts written on individual corporates and are traded on secundary financial markets. In the simplest case of **digital CDS**, the payoff is equal to:

$$
\begin{aligned}
&1\$, \quad \text{if the corporation is still alive at } t+h, \\
&0\$, \quad \text{otherwise.}
\end{aligned}
$$

Such CDS are regularly traded on the market for a variety of names. The price of such a CDS for maturity h is always smaller than the price of the riskfree zero-coupon bond with the same maturity. The ratio of the prices at t of the CDS and of the associated riskfree bond is a quantity $y_{i,t,h}$ between 0 and 1, which can be interpreted as the market price at t of default of the firm between t and $t+h$.

We expect that the common factor involved in individual default occurrences will also have an effect on the associated CDS prices. For this reason, some indexes of CDS prices analogues to an average:

$$\bar{y}_{n,t,h} = \frac{1}{n}\sum_{i=1}^{n} y_{i,t,h},$$

are regularly published. Examples are the **iTraxx indexes** covering the European and Asian markets and the **CDX indexes** for North-American markets. These indexes can be used as support for derivatives. For instance, a synthetic CDO tranche written on iTraxx, with time-to-maturity 1 will pay at $t+1$:

$$a(y_{1,t+1,h}, \ldots, y_{n,t+1,h}) = (\bar{y}_{n,t+1,h} - \alpha_1)^+ - (\bar{y}_{n,t+1,h} - \alpha_2)^+.$$

Even if the design of the payoff is similar for the CDO tranche written an a basket of loans, the two types of derivatives are different, since they are not written on the same type of individual risks, which are default occurrences and levels of CDS prices, respectively.

iv) Securitization in Insurance

Similar principles have been followed in Insurance to hedge longevity risk, that is, the uncertain general increase of human life. In practice the derivatives are written on observed mortality rates of individuals. They can be computed on a given subpopulation in a country, such as the generations of male born in US between 1960 and 1965. This is the case for the **longevity bonds**. They can also correspond to a portfolio of life insurance contracts securitized by an Insurance company. They are called **Mortality Linked Securities (MLS)**.

5.4 Derivative pricing

i) No-arbitrage and stochastic discount factor

The no-arbitrage condition is the impossibility to make a certain positive gain at some future date with an initial zero (or negative) investment (see Review B.2). The no-arbitrage condition is equivalent to the existence of a pricing operator, characterized by a **stochastic discount factor (sdf)** [Harrison, Kreps (1979)]. More precisely, let us consider an information set at date t, which includes the current and past values of variables observed by the investors. In our framework this information set will be:

$$J_t = (\underline{y_{1,t}}, \ldots, \underline{y_{n,t}}, \underline{f_{t-1}}). \tag{5.9}$$

Thus, the investors know the individual risks, but have an imperfect knowledge on the common factor. They know its past, but not its current value. Then, a stochastic discount factor for period $(t, t+1)$ is a positive function $m_{t,t+1}$, which depends on information J_{t+1}.

Under no-arbitrage, there exists a sdf such that the prices at date t of European derivative assets with payoff a_{t+h} at $t+h$ can be written as:

$$\pi_t(a, h) = E_t[m_{t,t+1} \ldots m_{t+h-1,t+h} a_{t+h}], \tag{5.10}$$

where E_t denotes the expectation conditional on the information J_t at time t.

Later on, we assume that the sdf depends on the information by means of f_t only, that is, $m_{t,t+1} = m(f_t)$.

ii) Pricing basket derivatives

Let us now consider an homogenous set (basket) of risks $y_{i,t}$, with $i = 1, \dots, n$, $t = 1, \dots T$, satisfying the assumptions recalled in the introduction of Chapter 5. A basket derivative pays at $t+h$ an amount $a(y_{1,t+h}, \dots, y_{n,t+h})$, say. Its price at date t is:

$$\pi_t(a, h) = E_t[m_{t,t+1} \dots m_{t+h-1,t+h} a(y_{1,t+h}, \dots, y_{n,t+h})]. \tag{5.11}$$

By the iterated expectation theorem and by using the assumptions on the state and measurement equations, the price can also be written as:

$$\pi_t(a, h) = E_t[m(f_t)\psi(f_t, a, h)], \tag{5.12}$$

where:

$$\psi(f_t, a, h) = E[m(f_{t+1}) \dots m(f_{t+h-1}) a(y_{1,t+h}, \dots, y_{n,t+h}) | \underline{f_t}, \underline{y_{1t}}, \dots, \underline{y_{n,t}}]. \tag{5.13}$$

Thus, the price of the initial basket derivative with time-to-maturity h is equal to the price of a virtual short term derivative written on f_t with payoff $\psi(f_t, a, h)$ at $t+1$. Function $m(f_t)\psi(f_t, a, h)$ corresponds to function $\alpha^*(f_t, h)$ in Proposition 5.5.

We have seen in Corollary 5.3 that the conditional distribution of f_t given J_t can be approximated at order $1/n$ by the Gaussian distribution with pdf given by:

$$\hat{\varphi}_{n,t}(f_t) = \frac{1}{\sigma_{n,t}} \phi \left(\frac{f_t - \mu_{n,t}}{\sigma_{n,t}} \right), \tag{5.14}$$

where ϕ is the pdf of the standard Gaussian distribution and:

$$\mu_{n,t} = \hat{f}_{nt} + \frac{1}{n}[I_{nt}^{-1} \frac{\partial \log g}{\partial f_t}(\hat{f}_{nt} | \hat{f}_{n,t-1}) + \frac{1}{2} I_{nt}^{-2} K_{nt}], \quad \sigma_{n,t}^2 = \frac{1}{n} I_{nt}^{-1}.$$

We deduce the following proposition:

Proposition 5.7: *The price of the basket derivative paying* $a(y_{1,t+h}, \ldots, y_{n,t+h})$ *at* $t+h$ *is such that:*

$$\pi_t(a,h) = \int m(f_t)\psi(f_t,a,h)\hat{\varphi}_{n,t}(f_t)df_t + o(1/n),$$

where function ψ *is defined in (5.13) and pdf* $\hat{\varphi}_{n,t}$ *is given in (5.14).*

Up to order 1/n, the basket derivative price can be approximated by a function of $\hat{f}_{nt}$, $\hat{f}_{n,t-1}$, I_{nt} and $K_{n,t}$ only. This approximated price does not require the knowledge of any past observation of the common factor. This is important for the two following reasons:

i) First, even if the investors observe the lagged factor values, the econometricians do not. Nevertheless, the latter ones can approximate the derivative price rather accurately by taking into account the cross-sectional information.

ii) Second, indirect observation on the values of the underlying factor could be deduced from the prices of highly traded derivatives written on the $y_{i,t+h}$. Since these factor values are not needed to compute the approximate derivative price, the approximate pricing formula in Proposition 5.7 can be used at the creation of a new derivative market to propose a coherent system of quotes for derivatives. In this situation, the sdf $m(.)$ is not a market correction for risk, but reflects the risk aversion and choices of the monopolistic firm, which is quoting first. The sdf has to be updated during the emergence of this derivative market to account for the adjustment of derivative prices due to demand and supply.

5.5 Derivatives Written on a Factor Proxy

i) The derivatives and their prices

As mentioned earlier, basket derivatives are usually introduced on financial markets as instruments to hedge the common risks. Since the common factor is not observed, they are usually written on a suitable proxy of this factor reflecting its risk dynamics. We derive below another approximate pricing formula when the derivatives are written on the cross-sectional maximum likelihood estimator of factor f_t [see Gagliardini, Gourieroux (2011) for the proof]. We focus on derivatives with exponential payoff, since these are the basis for the pricing of derivatives with more general payoff (see Section 5.6).

Proposition 5.8: *The true price at time t of the derivative with payoff $\exp(u\hat{f}_{n,t+h})$ at time $t+h$ is:*

$$\pi_{n,t}(u,h) = \int m(f_t)\psi_n(f_t,u,h)\hat{\varphi}_{n,t}(f_t)df_t + o(1/n),$$

where:

$$\psi_n(f_t,u,h) = E[m(f_{t+1})\dots m(f_{t+h-1})\exp[uf_{t+h} - \frac{u}{2n}I_{t+h}^{-2}\beta_{t+h} + \frac{u^2}{2n}I_{t+h}^{-1})|f_t],$$

and pdf $\hat{\varphi}_{n,t}$ is given in (5.14), with:

$$\begin{aligned} I_{t+h} &= E\left[\frac{-\partial^2 \log h(y_{i,t+h}|f_{t+h})}{\partial f^2}|f_{t+h}\right], \\ \beta_{t+h} &= Cov\left[\frac{\partial \log h(y_{i,t+h}|f_{t+h})}{\partial f}, \frac{\partial^2 \log h(y_{i,t+h}|f_{t+h})}{\partial f^2} + \left(\frac{\partial \log h(y_{i,t+h}|f_{t+h})}{\partial f}\right)^2 |f_{t+h}\right]. \end{aligned}$$

Compared to the general result in Proposition 5.7, in the framework of Proposition 5.8 we exploit the large size n of the basket to approximate function $\psi(f_t,u,h)$ by means of an expectation w.r.t. the factor path $\psi_n(f_t,u,h)$, up to order $o(1/n)$. This simplifies considerably the numerical calculation of the approximate derivative price.

A BDS is an example of a basket derivative whose payoff is written on the ML estimate of the systematic risk factor. Indeed, let us consider a short term BDS. The individual risks are measured by the 0-1 default occurrences, with Bernoulli distribution $\mathcal{B}(1,f_t)$. The cross-sectional ML estimator of f_t

is equal to the observed default frequency $\hat{f}_{n,t} = \bar{y}_{n,t}$ [see 5.2 iii)]. This is exactly the proxy of the factor used as support for BDS.

Let us now consider a CDO tranche on iTraxx. The underlying individual risks correspond to the implied default probabilities equal to the ratios of the CDS prices by the associated riskfree zero-coupon bonds. The individual risk corresponds to a real variable taking value between 0 and 1. A beta model with factor [see 5.2 v)] is a natural choice for describing these risks. Unfortunately, the CDO tranche is written on the average $\bar{y}_{n,t}$, which is not equal to the cross-sectional ML estimator.

The adjustment coefficients I_t and β_t in Proposition 5.8 are given in Table 5.1 for the models introduced in Section 5.2.

Table 5.1: The adjustment coefficients

Gaussian linear factor model	$I_t = 1/\sigma^2$	$\beta_t = 0$
Stochastic volatility model with factor	$I_t = 1/(2f_t^2)$	$\beta_t = 0$
Dichotomous qualitative model with factor	$I_t = 1/[f_t(1-f_t)]$	$\beta_t = 0$
Gamma model with factor	$I_t = \frac{d\psi}{ds}(f_t)$	$\beta_t = \frac{d^2\psi}{ds^2}(f_t)$
Beta model with factor	$I_t = \frac{d\psi_\alpha}{ds}(f_t)$	$\beta_t = \frac{d^2\psi_\alpha}{ds^2}(f_t)$

ii) SRF model for default correlation

Let us consider basket derivatives written on a default frequency $\hat{f}_{n,t} = \bar{y}_{n,t}$, where the individual risks $y_{i,t}$ are 0-1 variables with distribution $\mathcal{B}(1, f_t)$. From Table 5.1 we get $I_{t+1} = 1/[f_{t+1}(1-f_{t+1})]$ and $\beta_{t+1} = 0$. We deduce that the true price of the exponential derivative with short time-to-maturity $h = 1$ is:

$$\pi_{n,t}(u,1) = \int m(f_t) E(\exp[u f_{t+1} + \frac{u^2}{2n} f_{t+1}(1 - f_{t+1})] | f_t) \hat{\varphi}_{n,t}(f_t) df_t + o(1/n), \tag{5.15}$$

where $\hat{\varphi}_{nt}$ is the Gaussian pdf (5.14) with:

$$\mu_{n,t} = \hat{f}_{n,t} + \frac{1}{n}[\hat{f}_{n,t}(1-\hat{f}_{n,t}) \frac{\partial \log g}{\partial f_t}(\hat{f}_{n,t}|\hat{f}_{n,t-1}) + (1-2\hat{f}_{n,t})], \quad \sigma^2_{n,t} = \frac{\hat{f}_{n,t}[1-\hat{f}_{n,t}]}{n}.$$

One GA term appears by means of the term $\frac{u^2}{2n} f_{t+1}(1 - f_{t+1})$ in formula (5.15). It involves the uncertainty on the probability of default at date $t+1$, since $V[\hat{f}_{n,t+1}|f_{t+1}] = \frac{f_{t+1}(1-f_{t+1})}{n}$. An increase in this uncertainty, for instance, if n diminishes, implies an increase in the derivative price, that is, in the price of the corresponding insurance product.

5.6 Application to Approximate Pricing of BDS

In this section we present a numerical illustration for the approximate pricing of BDS [see Gagliardini, Gourieroux (2011)].

i) The risk factor model

The risk variables $y_{i,t}$ are binary default indicators for an homogenous portfolio of corporate loans. Their joint distribution is given by a dynamic version of the Merton (1974)-Vasicek (1991) Value of the Firm model [see Chapter 3.1 ii)]. We have $y_{i,t} = 1$, if $A_{i,t} < L_{i,t}$, and $y_{i,t} = 0$, otherwise, where $A_{i,t}$ and $L_{i,t}$ denote the firm asset and liability, respectively. The log asset/liability ratios follow a linear single risk factor model:

$$\log(A_{i,t}/L_{i,t}) = -\Phi^{-1}(PD) + \sqrt{\rho} F_t + \sqrt{1-\rho}\, u_{i,t}, \tag{5.16}$$

where shocks $u_{i,t}$ are $IIN(0,1)$ across firms and time dates, $PD \in (0,1)$ and $\rho \in (0,1)$. The systematic factor F_t follows a Gaussian autoregressive process:

$$F_t = \gamma F_{t-1} + \sqrt{1-\gamma^2}\varepsilon_t, \tag{5.17}$$

where $\varepsilon_t \sim IIN(0,1)$ and the autoregressive coefficient γ is such that $|\gamma| < 1$. The stationary distribution of F_t is standard Gaussian. Then, the parameterization in (5.16) is such that PD is the unconditional default probability of a firm, while ρ is the contemporaneous correlation between the asset/liability ratios of two firms.

The time unit is 1 year. We set an unconditional 1-year default probability equal to $PD = 0.04$. We consider three values for the asset correlation, that are $\rho = 0.01,\ 0.10,\ 0.30$. They cover the range of asset correlation values which are compatible with default correlation estimates reported in the literature [De Servigny, Renault (2002), Gagliardini, Gourieroux (2005)], as well as values suggested by Basel II regulation [BCBS (2001), (2003)]. The portfolio size is $n = 1000$.

Model (5.16)-(5.17) is such that the default indicators $y_{i,t}$, for i varying, are i.i.d. conditional on factor F_t, with Bernoulli distribution $\mathcal{B}(1, f_t)$, where the transformed factor f_t is the conditional default probability:

$$f_t = P[\log(A_{i,t}/L_{i,t}) < 0 | F_t] = \Phi\left(\frac{\Phi^{-1}(PD) - \sqrt{\rho}F_t}{\sqrt{1-\rho}}\right), \tag{5.18}$$

[see Example iii) in Section 5.2]. The transition density of Markov process (f_t) is deduced from (5.17) and (5.18):

$$g(f_t|f_{t-1}) = \frac{1}{\sqrt{1-\gamma^2}}\phi\left(\frac{F_t - \gamma F_{t-1}}{\sqrt{1-\gamma^2}}\right)\frac{\sqrt{1-\rho}}{\sqrt{\rho}}\frac{1}{\phi[\Phi^{-1}(f_t)]}, \tag{5.19}$$

where $F_t = \dfrac{\Phi^{-1}(PD) - \sqrt{1-\rho}\Phi^{-1}(f_t)}{\sqrt{\rho}}$. This transition pdf is displayed in Figure 5.2 for different values of the lagged factor when asset correlation is $\rho = 0.10$.

[Insert Figure 5.2: Transition pdf of transformed factor f_t]

The lagged value f_{t-1} has an impact on both the location and the variance of the distribution of f_t. Moreover, the transition pdf is right skewed.

ii) Approximate filtering distribution

In order to assess the accuracy of the Gaussian approximation in Proposition 5.1 and its Corollaries, let us now compare the distribution of factor f_t given the investor information J_t and the Gaussian approximate filtering distribution in Corollary 5.3. The pdf of f_t conditional on $J_t = (\underline{f_{t-1}}, \underline{y_{1,t}}, ..., \underline{y_{n,t}})$ is given by:

$$\begin{aligned} g(f_t|J_t) &= \frac{g(\underline{f_t}, \underline{y_{1,t}}, ..., \underline{y_{n,t}})}{\int g(\underline{f_t}, \underline{y_{1,t}}, ..., \underline{y_{n,t}}) df_t} = \frac{\prod_{i=1}^{n} h(y_{i,t}|f_t) g(f_t|f_{t-1})}{\int \prod_{i=1}^{n} h(y_{i,t}|f_t) g(f_t|f_{t-1}) df_t} \\ &= \frac{f_t^{n\bar{y}_{n,t}} (1-f_t)^{n(1-\bar{y}_{n,t})} g(f_t|f_{t-1})}{\int f_t^{n\bar{y}_{n,t}} (1-f_t)^{n(1-\bar{y}_{n,t})} g(f_t|f_{t-1}) df_t}, \end{aligned} \tag{5.20}$$

and depends on the conditioning information J_t by means of current default frequency $\bar{y}_{n,t}$ and lagged factor f_{t-1} only. The Gaussian approximation of the filtering ditribution is obtained from Corollary 5.3 with factor approximation equal to the default frequency $\hat{f}_{n,t} = \bar{y}_{n,t}$, statistics $I_{n,t}$, $K_{n,t}$ given in (5.7) [see Section 5.2 iii)], and partial derivative of the factor log-transition density $\frac{\partial \log g}{\partial f_t}(\hat{f}_{n,t}|\hat{f}_{n,t-1})$ computed from (5.19). The approximate filtering distribution depends on default history by means of current and lagged default frequencies $\hat{f}_{n,t}$ and $\hat{f}_{n,t-1}$ only.

In Figure 5.3 we display the predictive distribution of f_t given investor information J_t, for $\hat{f}_{n,t} = 0.04$ and different values of the lagged factor f_{t-1}. In Figure 5.4 we display the approximate filtering distribution of f_t, for $\hat{f}_{n,t} = 0.04$ and different values of the lagged default frequancy $\hat{f}_{n,t-1}$. Asset correlation is $\rho = 0.10$.

[Insert Figure 5.3: Predictive distribution of factor f_t given the investor information]

[Insert Figure 5.4: Approximate filtering distribution of factor f_t]

By comparing Figures 5.2 and 5.3, it is seen that the default history is very informative for the distribution of the unobservable factor. Indeed, by including default frequency $\hat{f}_{n,t}$ in the conditioning set, the distribution of f_t given f_{t-1} is less dispersed, and much closer to a Gaussian distribution. Moreover, this distribution is rather insensitive to the lagged factor value f_{t-1}, as explained by Corollary 5.2. Similarly, the Gaussian approximation of the filtering distribution in Figure 5.4 is quite independent of the lagged default frequency $\hat{f}_{n,t-1}$. Finally, by comparing Figures 5.3 and 5.4 we deduce that the Gaussian approximation of the filtering distribution is rather accurate.

iii) Approximate pricing of BDS

Let us now consider the approximate pricing of short-term BDS. For expository purpose, we consider the sdf $m_{t,t+1} = 1$, that is, we set the risk-free rate and the risk premium for systematic risk equal to zero. The derivative payoff is $a(y_{1,t+1}, ..., y_{n,t+1}) = \mathbb{1}_{\hat{f}_{n,t+1} > \alpha}$, with $\alpha \in (0,1)$ [see Section 5.3 i)]. By using the Fourier Transform Inversion formula [see e.g. Proposition 2 in Duffie, Pan, Singleton (2000)], it is possible to write the price $\pi_t(\alpha, 1)$ of such a derivative as an integral transform of the prices of derivatives with exponential payoff. More precisely, we have:

$$\pi_t(\alpha, 1) = \frac{1}{2} + \frac{1}{\pi} \int_0^\infty \frac{Im\left[\tilde{\pi}_t(iv, 1) \exp\left(-iv\alpha_t\right)\right]}{v} dv, \tag{5.21}$$

where i is the imaginary unit, Im denotes the imaginary part of a complex number, and:

$$\tilde{\pi}_t(u, 1) = E\left[\exp(u\hat{f}_{n,t+1}) | J_t\right]. \tag{5.22}$$

We use equation (5.12) and Proposition 5.7 to derive the true and approximate prices of derivatives with exponential payoff, and then apply transformation (5.21) to get the true and approximated prices of the BDS. The advantage of this approach by Fourier transform inversion is that function ψ in (5.13) can be computed in closed form for exponential derivatives, up to an expectation w.r.t. the factor value. More precisely, by the iterated expectation theorem we have:

$$\begin{aligned}\psi(f_t, u, 1) &= E\left[\exp(u\hat{f}_{n,t+1})|\underline{f_t}, \underline{y_{1,t}}, ..., \underline{y_{n,t}}\right] \\ &= E\left[E\left[\exp(u\hat{f}_{n,t+1})|\underline{f_{t+1}}, \underline{y_{1,t}}, ..., \underline{y_{n,t}}\right] |\underline{f_t}, \underline{y_{1,t}}, ..., \underline{y_{n,t}}\right].\end{aligned}$$

Then, by using that the risks $y_{i,t+1}$ are iid $\mathcal{B}(1, f_{t+1})$ given $\underline{f_{t+1}}, \underline{y_{1,t}}, ..., \underline{y_{n,t}}$, we get:

$$\begin{aligned}\psi(f_t, u, 1) &= E\left[\left(1 + (e^{u/n} - 1)f_{t+1}\right)^n |\underline{f_t}, \underline{y_{1,t}}, ..., \underline{y_{n,t}}\right] \\ &= E\left[\left(1 + (e^{u/n} - 1)f_{t+1}\right)^n |f_t\right].\end{aligned}$$

Thus, the true and approximated prices of the exponential derivatives are $\tilde{\pi}_t(u, 1) = \int \psi(f_t, u, 1)g(f_t|J_t)df_t$ and $\tilde{\pi}_{n,t}(u, 1) = \int \psi(f_t, u, 1)\hat{\phi}_{n,t}(f_t)df_t$, respectively, where the pdf $g(\cdot|J_t)$ given the investor information and the approximate filtering distribution $\hat{\phi}_{n,t}(\cdot)$ are derived in Section ii). The true price depends on the available information by means of $\hat{f}_{n,t}$ and f_{t-1}, while the approximate price depends on the default history by means of $\hat{f}_{n,t}$ and $\hat{f}_{n,t-1}$. The expectations w.r.t. f_{t+1} in function ψ, and w.r.t. f_t in the true and approximated prices, can be computed by Monte-Carlo integration. The integral in (5.21) can be computed by numerical integration.

In Figures 5.5 and 5.6 we display the true and approximate BDS price for time-to-maturity $h = 1$ year, respectively, as a function of threshold α, and for different values of default correlation. In Figure 5.5 the available information is such that $\hat{f}_{n,t} = f_{t-1} = 0.04$, and in Figure 5.6 the default history is such that $\hat{f}_{n,t} = \hat{f}_{n,t-1} = 0.04$.

[Insert Figure 5.5: True price of the BDS]

[Insert Figure 5.6: Approximate price of the BDS]

The true BDS price is clearly a decreasing function of the threshold α. Its pattern corresponds to the (risk-neutral) conditional survivor function of the future default frequency $\hat{f}_{n,t+1}$ given the available information. For small values of the asset correlation, the BDS price is close to 1 for α smaller than the current default frequency $\hat{f}_{n,t} = 0.04$, and close to 0 for α larger than $\hat{f}_{n,t} = 0.04$. In the latter case the BDS price corresponds to the market price of a rare joint default event. The default correlation parameter ρ has a significant impact on the BDS price. On Figure 5.5 it is seen that an increase of the asset correlation ρ implies an increase of the BDS price for α above $\hat{f}_{n,t}$, and a decrease for α below $\hat{f}_{n,t}$. This is due to the positive effect of asset correlation ρ on the conditional variance of $\hat{f}_{n,t+1}$ given the available information. By comparing Figures 5.5 and 5.6 we deduce that the approximation of the BDS price provided by Proposition 5.7 is rather accurate.

5.7 Summary

In factor models the prediction and filtering formulas involve large-dimensional integrals. However, for large panels, these formulas can be approximated under closed form at order $1/n$, where n is the cross-sectional dimension. These approximations correspond to the standard prediction and filtering formulas applied to an appropriately linearized state space model. These approximated prediction formulas can be applied to compute at order $1/n$ the prices of derivatives on a basket of individual risks.

5.8 Appendix: Approximation of the Filtering Distribution

i) Proof of Proposition 5.1

Let us first derive an approximation for the conditional Laplace transform of f_t given $\underline{y_{1,t}}$, ...,$\underline{y_{n,t}}$ and $\underline{f_{t-1}}$:

$$\mathcal{L}_{nt}(u) = E\left[\exp\left(uf_t\right)|\underline{y_{1,t}}, ..., \underline{y_{n,t}}, \underline{f_{t-1}}\right] = \frac{\displaystyle\int e^{uf_t} g(f_t|f_{t-1}) \prod_{i=1}^{n} h\left(y_{i,t}|f_t\right) df_t}{\displaystyle\int g(f_t|f_{t-1}) \prod_{i=1}^{n} h\left(y_{i,t}|f_t\right) df_t}, \tag{5.23}$$

which depends only on $y_{1,t}, ..., y_{nt}$ and f_{t-1}.

Let us expand the micro-density around $\hat{f}_{nt}$:

$$\begin{aligned}
\sum_{i=1}^{n} \log h\left(y_{i,t}|f_t\right) &= \sum_{i=1}^{n} \log h\left(y_{i,t}|\hat{f}_{nt}\right) \\
&+ \frac{1}{2}\frac{1}{n}\sum_{i=1}^{n} \frac{\partial^2 \log h}{\partial f_t^2}\left(y_{i,t}|\hat{f}_{nt}\right)\left[\sqrt{n}\left(f_t - \hat{f}_{nt}\right)\right]^2 \\
&+ \frac{1}{6\sqrt{n}}\frac{1}{n}\sum_{i=1}^{n} \frac{\partial^3 \log h}{\partial f_t^3}\left(y_{i,t}|\hat{f}_{nt}\right)\left[\sqrt{n}\left(f_t - \hat{f}_{nt}\right)\right]^3 \\
&+ \frac{1}{24n}\frac{1}{n}\sum_{i=1}^{n} \frac{\partial^4 \log h}{\partial f_t^4}\left(y_{i,t}|\hat{f}_{nt}\right)\left[\sqrt{n}\left(f_t - \hat{f}_{nt}\right)\right]^4 + o\left(1/n\right).
\end{aligned}$$

Let us introduce the change of variable:

$$X = I_{nt}^{1/2}\sqrt{n}\left(f_t - \hat{f}_{nt}\right) \quad \Longleftrightarrow \quad f_t = \hat{f}_{nt} + \frac{1}{\sqrt{n}} I_{nt}^{-1/2} X.$$

Then, we have:

$$\sum_{i=1}^{n} \log h\left(y_{i,t}|f_t\right) = \sum_{i=1}^{n} \log h\left(y_{i,t}|\hat{f}_{nt}\right) - \frac{1}{2}X^2 + \frac{1}{6\sqrt{n}} J_{nt} X^3 + \frac{1}{24n} Q_{nt} X^4 + o(1/n),$$

where:

$$J_{nt} = I_{nt}^{-3/2} K_{nt} \quad \text{and} \quad Q_{nt} = I_{nt}^{-2} \frac{1}{n} \sum_{i=1}^{n} \frac{\partial^4 \log h}{\partial f_t^4} \left(y_{i,t} | \hat{f}_{nt} \right).$$

Thus:

$$\begin{aligned}
\prod_{i=1}^{n} h\left(y_{i,t}|f_t\right) &= \prod_{i=1}^{n} h\left(y_{i,t}|\hat{f}_{nt}\right) \exp\left(-\frac{1}{2}X^2\right) \exp\left(\frac{1}{6\sqrt{n}} J_{nt} X^3 + \frac{1}{24n} Q_{nt} X^4 + o(1/n)\right) \\
&= \prod_{i=1}^{n} h\left(y_{i,t}|\hat{f}_{nt}\right) \exp\left(-\frac{1}{2}X^2\right) \\
&\quad \left[1 + \frac{1}{6\sqrt{n}} J_{nt} X^3 + \frac{1}{24n} Q_{nt} X^4 + \frac{1}{72n} J_{nt}^2 X^6 + o(1/n)\right]. \qquad (5.24)
\end{aligned}$$

Similarly, we have an expansion for $\log g(f_t|f_{t-1})$ as:

$$\begin{aligned}
\log g(f_t|f_{t-1}) &= \log g\left(\hat{f}_{nt} + \frac{1}{\sqrt{n}} I_{nt}^{-1/2} X | f_{t-1}\right) \\
&= \log g\left(\hat{f}_{nt}|f_{t-1}\right) + \frac{1}{\sqrt{n}} I_{nt}^{-1/2} A_{nt} X + \frac{1}{2n} I_{nt}^{-1} B_{nt} X^2 + o(1/n),
\end{aligned}$$

where:

$$A_{nt} = \frac{\partial \log g}{\partial f_t} \left(\hat{f}_{nt}|f_{t-1}\right) \quad \text{and} \quad B_{nt} = \frac{\partial^2 \log g}{\partial f_t^2} \left(\hat{f}_{nt}|f_{t-1}\right).$$

Thus:

$$\begin{aligned}
g(f_t|f_{t-1}) &= g\left(\hat{f}_{nt}|f_{t-1}\right) \exp\left(\frac{1}{\sqrt{n}} I_{nt}^{-1/2} A_{nt} X + \frac{1}{2n} I_{nt}^{-1} B_{nt} X^2 + o(1/n)\right) \\
&= g\left(\hat{f}_{nt}|f_{t-1}\right) \left[1 + \frac{1}{\sqrt{n}} I_{nt}^{-1/2} A_{nt} X + \frac{1}{2n} I_{nt}^{-1} B_{nt} X^2 + \frac{1}{2n} I_{nt}^{-1} A_{nt}^2 X^2 + o(1/n)\right].
\end{aligned} \qquad (5.25)$$

Finally, we have an expansion for $\exp(uf_t)$:

$$\begin{aligned}
\exp(uf_t) &= \exp\left(u\hat{f}_{nt}\right) \exp\left(\frac{u}{\sqrt{n}} I_{nt}^{-1/2} X\right) \\
&= \exp\left(u\hat{f}_{nt}\right) \left[1 + \frac{u}{\sqrt{n}} I_{nt}^{-1/2} X + \frac{u^2}{2n} I_{nt}^{-1} X^2 + o(1/n)\right].
\end{aligned} \qquad (5.26)$$

Let us now substitute expansions (5.24)-(5.26) into the numerator in equation (5.23) (the denominator is obtained by setting $u = 0$). We have:

$$\int e^{uf_t} g(f_t|f_{t-1}) \prod_{i=1}^{n} h\left(y_{i,t}|f_t\right) df_t = e^{u\hat{f}_{nt}} \prod_{i=1}^{n} h\left(y_{i,t}|\hat{f}_{nt}\right) g\left(\hat{f}_{nt}|f_{t-1}\right)$$

$$E_X\left[\left(1+\frac{u}{\sqrt{n}}I_{nt}^{-1/2}X+\frac{u^2}{2n}I_{nt}^{-1}X^2+o(1/n)\right)\right.$$
$$\left(1+\frac{1}{\sqrt{n}}I_{nt}^{-1/2}A_{nt}X+\frac{1}{2n}I_{nt}^{-1}\left(B_{nt}+A_{nt}^2\right)X^2+o(1/n)\right)$$
$$\left.\left(1+\frac{1}{6\sqrt{n}}J_{nt}X^3+\frac{1}{24n}Q_{nt}X^4+\frac{1}{72n}J_{nt}^2X^6+o(1/n)\right)\right],$$

where the expectation E_X is w.r.t. the standard normal variable X. Since odd power moments of X are equal to zero, the terms of order $1/\sqrt{n}$ [and similarly the terms of order $1/\left(n\sqrt{n}\right)$, if the expansion is considered up to order $1/n^2$] cancel and the expectation is equal to:

$$1+\frac{u}{n}\left[I_{nt}^{-1}A_{nt}+\frac{1}{2}I_{nt}^{-1/2}J_{nt}\right]+\frac{1}{2n}u^2I_{nt}^{-1}+\Lambda_{nt}+O(1/n^2),$$

where:

$$\Lambda_{nt}=\frac{1}{2n}I_{nt}^{-1}\left(B_{nt}+A_{nt}^2\right)+\frac{1}{2n}I_{nt}^{-1/2}J_{nt}A_{nt}+\frac{1}{8n}Q_{nt}+\frac{1}{72n}J_{nt}^2E\left[X^6\right],$$

is independent of u. Thus, we deduce:

$$\begin{aligned}
&\mathcal{L}_{nt}(u)\\
&= e^{u\hat{f}_{nt}}\frac{1+\frac{u}{n}\left(I_{nt}^{-1}A_{nt}+\frac{1}{2}I_{nt}^{-1/2}J_{nt}\right)+\frac{1}{2n}u^2I_{nt}^{-1}+\Lambda_{nt}+O(1/n^2)}{1+\Lambda_{nt}+O(1/n^2)}\\
&= e^{u\hat{f}_{nt}}\left(1+\frac{u}{n}\left(I_{nt}^{-1}A_{nt}+\frac{1}{2}I_{nt}^{-1/2}J_{nt}\right)+\frac{u^2}{2n}I_{nt}^{-1}+\Lambda_{nt}+O(\frac{1}{n^2})\right)\left(1-\Lambda_{nt}+O(\frac{1}{n^2})\right)\\
&= e^{u\hat{f}_{nt}}\left[1+\frac{u}{n}\left(I_{nt}^{-1}A_{nt}+\frac{1}{2}I_{nt}^{-1/2}J_{nt}\right)+\frac{u^2}{2n}I_{nt}^{-1}+O(\frac{1}{n^2})\right].
\end{aligned}$$

By definition of J_{nt} and A_{nt}, we conclude:

$$\mathcal{L}_{nt}(u)=e^{u\hat{f}_{nt}}\left\{1+\frac{u}{n}\left(I_{nt}^{-1}\frac{\partial\log g}{\partial f_t}\left(\hat{f}_{nt}|f_{t-1}\right)+\frac{1}{2}I_{nt}^{-2}K_{nt}\right)+\frac{1}{2n}u^2I_{nt}^{-1}\right\}+O(1/n^2),$$

and

$$\mathcal{L}_{nt}(u) = \exp\left\{ u\hat{f}_{nt} + \frac{u}{n}\left(I_{nt}^{-1}\frac{\partial \log g}{\partial f_t}\left(\hat{f}_{nt}|f_{t-1}\right) + \frac{1}{2}I_{nt}^{-2}K_{nt}\right) + \frac{1}{2n}u^2 I_{nt}^{-1} + O(1/n^2)\right\}.$$

Another approximation valid at order $1/n$ can be obtained by replacing f_{t-1} by $\hat{f}_{n,t-1}$. We have:

$$\mathcal{L}_{nt}(u) = \exp\left\{ u\hat{f}_{nt} + \frac{u}{n}\left(I_{nt}^{-1}\frac{\partial \log g}{\partial f_t}\left(\hat{f}_{nt}|\hat{f}_{n,t-1}\right) + \frac{1}{2}I_{nt}^{-2}K_{nt}\right) + \frac{1}{2n}u^2 I_{nt}^{-1} + o(1/n)\right\}.$$

Then, Proposition 5.1 follows.

ii) Proof of Corollary 5.4

Let us expand function a at second-order around $\hat{f}_{nt}$:

$$a\left(f_t\right) = a\left(\hat{f}_{nt}\right) + \frac{da}{df}\left(\hat{f}_{nt}\right)\left(f_t - \hat{f}_{nt}\right) + \frac{1}{2}\frac{d^2a}{df^2}\left(\hat{f}_{nt}\right)\left(f_t - \hat{f}_{nt}\right)^2 + o\left(\left(f_t - \hat{f}_{nt}\right)^2\right).$$

Then, by computing the conditional expectation w.r.t. the Gaussian density for f_t given in Corollary 5.3, Corollary 5.4 follows.

iii) Approximate linear state space model and Kalman filter

Let us prove that the filtering distribution obtained by applying the Kalman filter on the linear state space model (5.3)-(5.4) equals the Gaussian distribution in Corollary 5.3 up to order $o(1/n)$. For expository purpose, we set $\mu = 0$ (the proof for $\mu \neq 0$ is similar).

From the Kalman filter (see Review A.5), the distribution of f_t given the history $\xi_{n,t}$, $\xi_{n,t-1}$, ... is Gaussian, with mean $\hat{f}_{t|t}$ and variance $\hat{\Sigma}_{t|t}$ satisfying recursive equations. To write these equations, let $\hat{f}_{t|t-1}$ and $\hat{\Sigma}_{t|t-1}$ denote the conditional mean and variance of factor f_t given the lagged information $\xi_{n,t-1}$, $\xi_{n,t-2}$, ... Then, from Review A.5 we have:

$$\begin{aligned} \hat{f}_{t|t} &= \hat{f}_{t|t-1} + K_{t|t}(\xi_{n,t} - \hat{f}_{t|t-1}) \\ &= \gamma(1 - K_{t|t})\hat{f}_{t-1|t-1} + K_{t|t}\xi_{n,t}, \end{aligned} \qquad (5.27)$$

and:

$$\Sigma_{t|t} = (1 - K_{t|t})\Sigma_{t|t-1}, \qquad (5.28)$$

where the Kalman gain $K_{t|t}$ is such that:

$$K_{t|t} = \frac{\Sigma_{t|t-1}}{\Sigma_{t|t-1} + \frac{1}{n} I_{n,t}^{-1}}, \tag{5.29}$$

and:

$$\Sigma_{t|t-1} = \gamma^2 \Sigma_{t-1|t-1} + \eta^2. \tag{5.30}$$

From equations (5.28)-(5.30) we deduce:

$$\begin{aligned} \Sigma_{t|t} &= \frac{1}{n} I_{n,t}^{-1} + o(1/n), \qquad (5.31) \\ \Sigma_{t|t-1} &= \eta^2 + O(1/n), \quad K_{t|t} = 1 - \frac{1}{n\eta^2} I_{n,t}^{-1} + o(1/n). \end{aligned}$$

Then, from equation (5.27) we get:

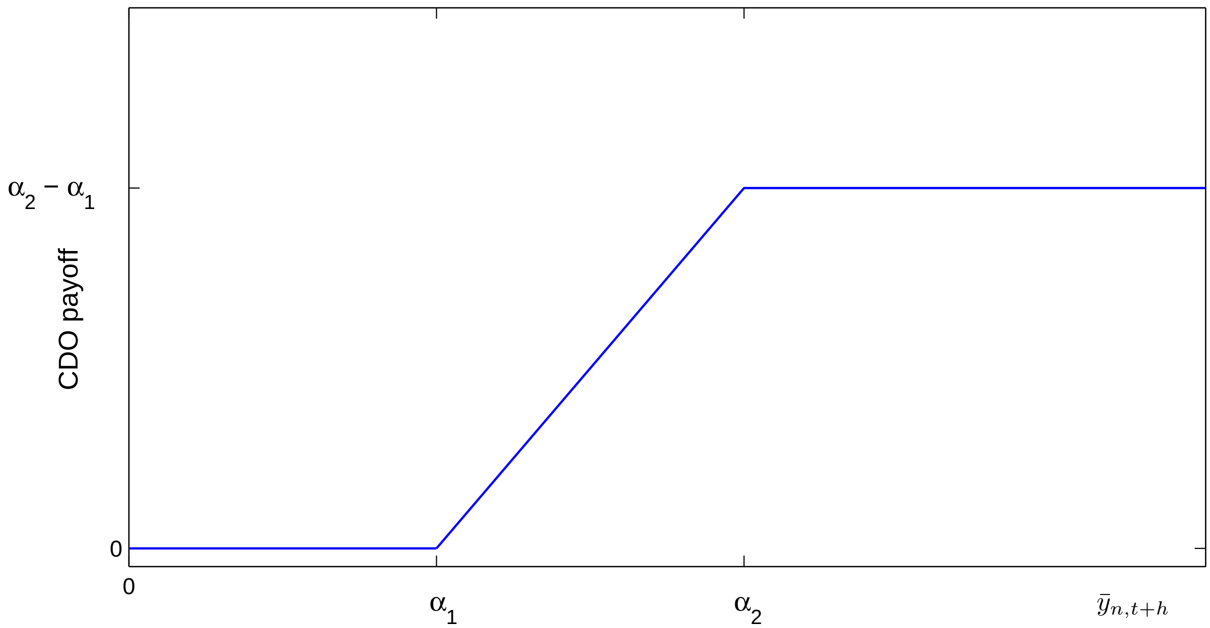
$$\begin{aligned} \hat{f}_{t|t} &= \frac{\gamma}{n\eta^2} I_{n,t}^{-1} \hat{f}_{t-1|t-1} + \xi_{n,t} - \frac{1}{n\eta^2} I_{n,t}^{-1} \xi_{n,t} + o(1/n) \\ &= \hat{f}_{n,t} + \frac{1}{2n} I_{n,t}^{-2} K_{n,t} - \frac{1}{n\eta^2} I_{n,t}^{-1} (\hat{f}_{n,t} - \gamma \hat{f}_{t-1|t-1}) + o(1/n). \end{aligned}$$

We deduce:

$$\hat{f}_{t|t} = \hat{f}_{n,t} + \frac{1}{2n} I_{n,t}^{-2} K_{n,t} - \frac{1}{n\eta^2} I_{n,t}^{-1} (\hat{f}_{n,t} - \gamma \hat{f}_{n,t-1}) + o(1/n). \tag{5.32}$$

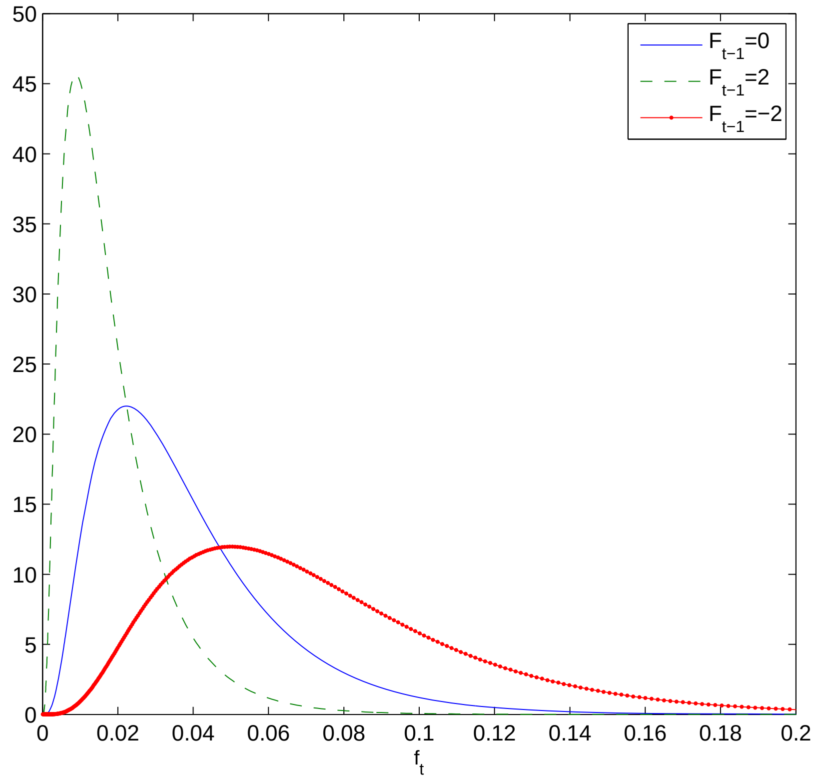
Then, by using that $\frac{\partial \log g}{\partial f_t}(f_t|f_{t-1}) = -\frac{f_t - \gamma f_{t-1}}{\eta^2}$, from equations (5.31) and (5.32) the conclusion follows.

Figure 5.1: Payoff of a CDO tranche.



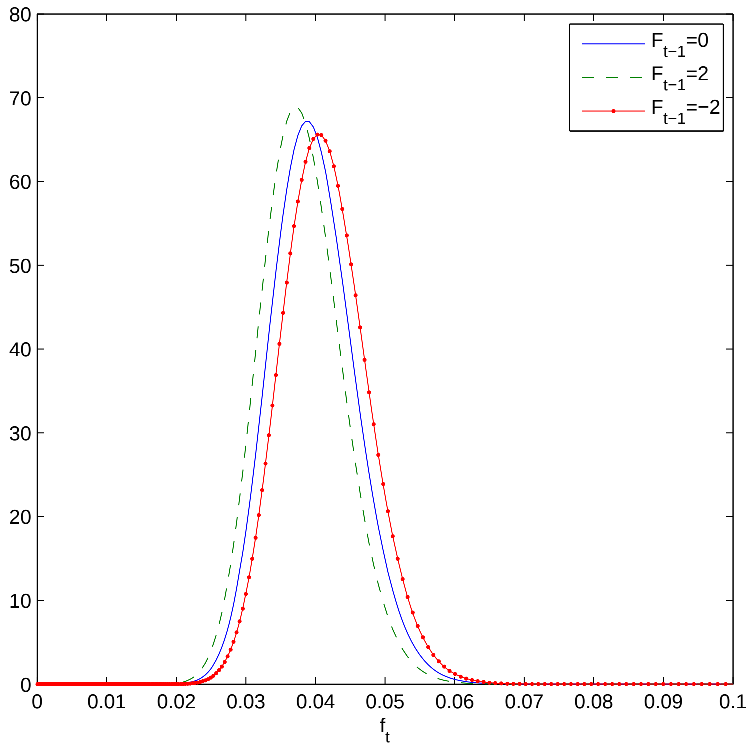
The figure displays the payoff of a CDO tranche as a function of the default frequency $\bar{y}_{t+h}$ at maturity. Thresholds α_1 and α_2 are the attachment and detachment points, respectively.

Figure 5.2: Transition pdf of transformed factor f_t.



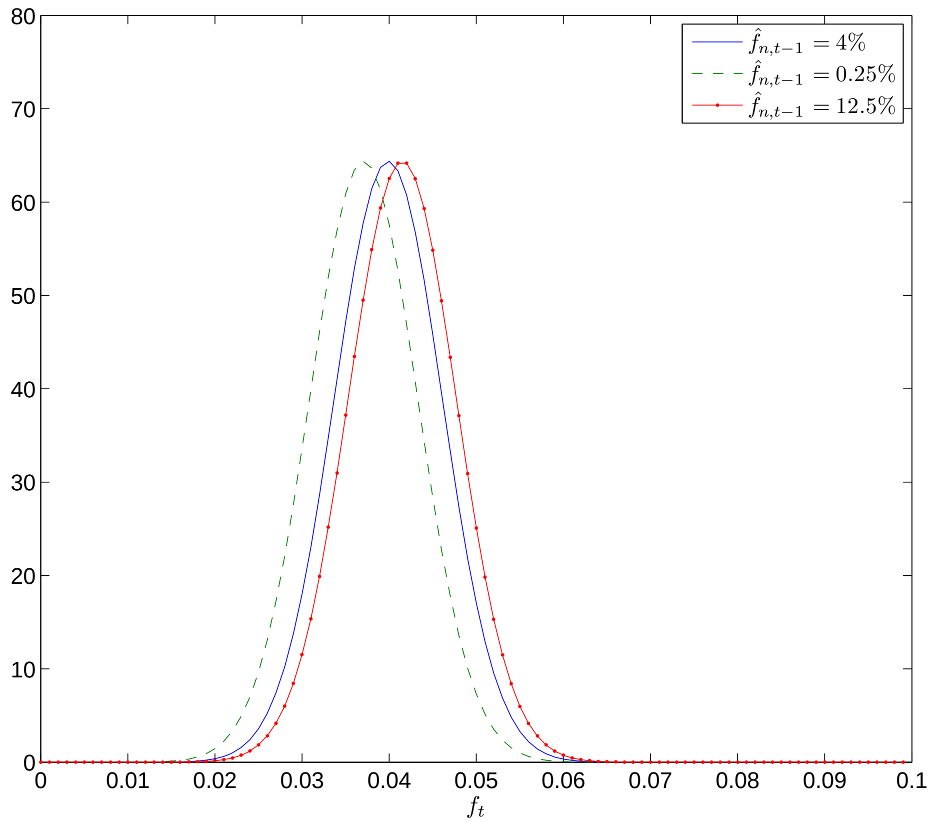
The Figure plots the conditional distribution of factor f_t given lagged value f_{t-1}, for different values of f_{t-1}. The conditioning values f_{t-1} are given in terms of their corresponding Gaussian factor values F_{t-1}; they are $F_{t-1} = 0$, $F_{t-1} = 2$, $F_{t-1} = -2$, respectively. Asset correlation is $\rho = 0.10$.

Figure 5.3: Predictive distribution of f_t given the investor information.



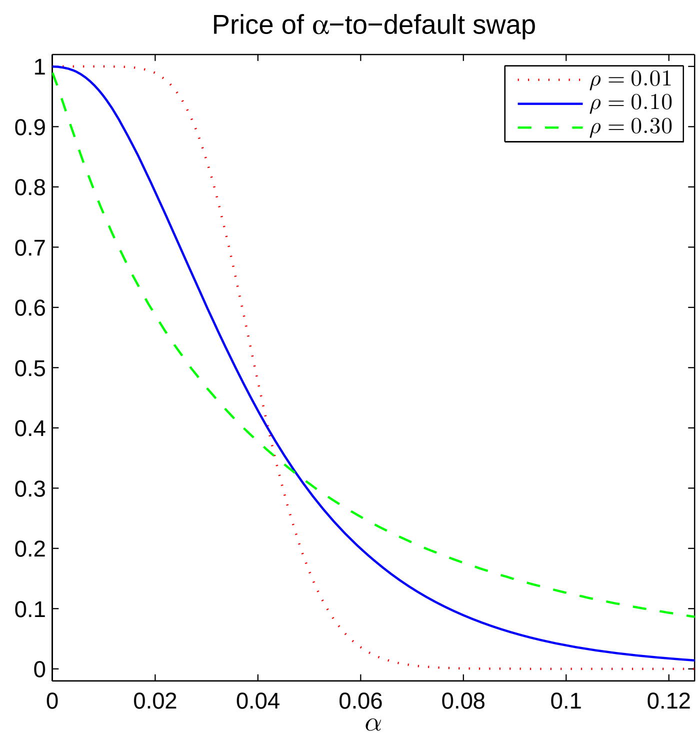
The Figure plots the conditional distribution of factor f_t given investor information J_t, such that $\hat{f}_{n,t} = 0.04$, and for different values of f_{t-1}. The conditioning values of f_{t-1} are given in terms of their corresponding Gaussian factor values F_{t-1}; they are $F_{t-1} = 0$, $F_{t-1} = 2$, and $F_{t-1} = -2$ respectively. Asset correlation is $\rho = 0.10$.

Figure 5.4: Approximate filtering distribution of f_t.



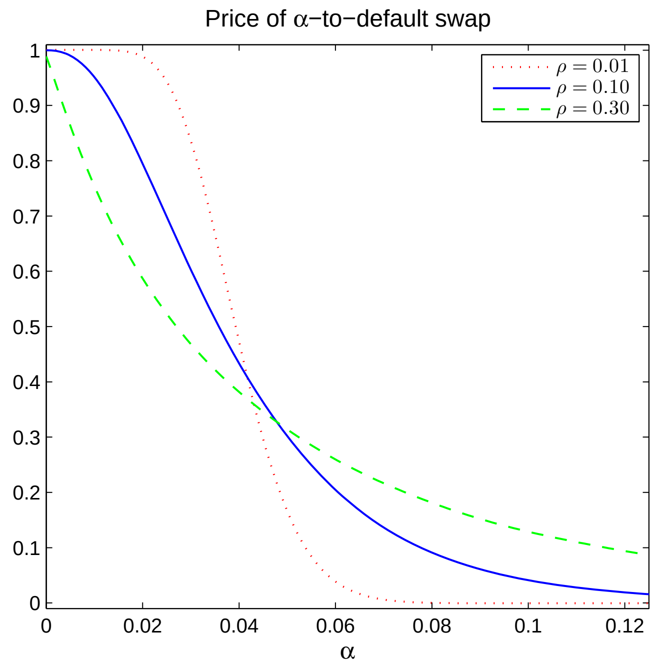
The Figure plots the approximate distribution of f_t given past default history, such that $\hat{f}_{n,t} = 0.04$, and for different values of $\hat{f}_{n,t-1}$. Asset correlation is $\rho = 0.10$.

Figure 5.5: True price of the BDS.



The Figure plots the price of the BDS at time-to-maturity 1 year as a function of threshold α, for three different values of asset correlation $\rho = 0.01$ (dotted line), $\rho = 0.10$ (solid line) and $\rho = 0.30$ (dashed line). The available information is such that $\hat{f}_{n,t} = f_{t-1} = 0.04$.

Figure 5.6: Approximate price of the BDS.



The Figure plots the approximate price of the BDS at time-to-maturity 1 year as a function of threshold α, for three different values of asset correlation $\rho = 0.01$ (dotted line), $\rho = 0.10$ (solid line) and $\rho = 0.30$ (dashed line). The past default history is such that $\hat{f}_{n,t} = \hat{f}_{n,t-1} = 0.04$.

R E F E R E N C E S

Basel Committee on Banking Supervision (2001): "The New Basel Capital Accord", Consultative Document of the Bank for International Settlements, April 2001, Part 2: Pillar 1.

Basel Committee on Banking Supervision (2003): "The New Basel Capital Accord", Consultative Document of the Bank for International Settlements, April 2003, Part 3: The Second Pillar.

De Servigny, A., and O., Renault (2002): "Default Correlation: Empirical Evidence", S&P, October.

Duffie, D., Pan, J., and K., Singleton (2000): "Transform Analysis and Asset Pricing for Affine Jump Diffusions", Econometrica, 68, 1343-1376.

Gagliardini, P., and C., Gouriéroux (2005): "Stochastic Migration Models with Application to Corporate Risk", Journal of Financial Econometrics, 3, 188-226.

Gagliardini, P., and C., Gourieroux (2010): "Efficiency in Large Dynamic Panel Models with Common Factor", CREST DP.

Gagliardini, P., and C., Gourieroux (2011): "Approximate Derivative Pricing in Large Classes of Homogeneous Assets with Systematic Risk", forthcoming in the Journal of Financial Econometrics.

Gourieroux, C., and A., Monfort (2008): "Quadratic Stochastic Intensity and Prospective Mortality Tables", Insurance: Mathematics and Economics, 43, 174-184.

Ito, K. (1951): On Stochastic Differential Equations, Memoirs of the American Mathematical Society, 4, 1-51.

Harrison, J., and D., Kreps (1979): "Martingales and Arbitrage in Multiperiod Securities Markets", Journal of Economic Theory, 20, 381-408.

Lamb, R., Perraudin, W., and A., Van Landschoot (2008): "Dynamic

Pricing of Synthetic Collaterized Debt Obligation", Imperial College W.P.

Laurent, J.P., and J., Gregory (2005): Basket Default Swaps, CDO's and Factor Copulas", Journal of Risk, 7, 103-122.

Merton, R. (1974): "On the Pricing of Corporate Debt: The Risk Structure of Interest Rates", Journal of Finance, 29, 449-470.

Vasicek, O. (1991): "Limiting Loan Loss Probability Distribution", KMV Technical Report.

Chapter 6

Granularity for Risk Measures

The current interest in risk measures is explained by the recent changes in regulation in the Finance and Insurance industries. New measures of risk have been introduced and are commonly used for risk management and risk control. In particular, they are the basis for determining the regulatory capital required to hedge the risk of a portfolio, or of a business line in a balance sheet.

The main risk measures, that are, the **Value-at-Risk (VaR)**, the **Expected Shortfall (ES)** or TailVaR, and the **Distortion Risk Measures (DRM)** are introduced in Section 6.1. Section 6.2 deals with the local analysis of risk measures, that is, their sensitivity to shocks on the distribution of the portfolio value. This local analysis is used in Section 6.3 to determine the granularity adjustment on theoretical risk measures for large homogenous portfolios in a static factor model. The extension to dynamic factor models is discussed in Section 6.4, where we consider how to account for the nonobservability of the current and lagged factor values. We finally consider in Section 6.5 the computation of risk measures for large portfolios of derivative assets written on a factor proxy (see also Chapter 5).

6.1 Risk Measures

Let us consider a given portfolio of assets. This portfolio can include stocks, corporate bonds, consumer loans, mortgages, or life insurance contracts. At date t, the value W_t of this portfolio is known, but its future value W_{t+h} at horizon h is unknown. This uncertainty is summarized in the **Profit and Loss (P&L)** distribution, which gives the conditional distribution of W_{t+h} given the information available at time t. To hedge this uncertainty some reserves R are introduced. With these reserves, which receive a zero return, the total value of the portfolio at date $t+h$ becomes $W_{t+h} + R$.

i) Value-at-Risk

Let us fix a probability of loss α, where $\alpha = 1$, 5, or 10%, say. The **reserve level** can be chosen such that:

$$P_t(W_{t+h} + R < 0) = \alpha, \tag{6.1}$$

where P_t denotes the conditional P&L distribution. By solving equation (6.1), we see that the corresponding reserve level is the opposite of the α-quantile of the P&L distribution. This level of reserve:

$$R = R(t, h, \alpha), \tag{6.2}$$

depends on date t, in particular on the information available at this date, on horizon h (there is a term structure of risk and a term structure of reserve), and on the loss probability α. This reserve $R(t, h, \alpha)$ is a decreasing function of α.

When the assets in the portfolio are stocks traded on the market, the portfolio value W_t generally features a nonstationary evolution, which can make difficult the determination of R. To circumvent this technical difficulty, it has been proposed to introduce the Value-at-Risk defined by:

$$VaR(t, h, \alpha) = R(t, h, \alpha) + W_t. \tag{6.3}$$

Thus, the VaR is characterized by:

$$P_t(W_{t+h} - W_t < -VaR(t, h, \alpha)] = \alpha, \tag{6.4}$$

and is the opposite of the conditional α-quantile of the distribution of change in portfolio value (see Figure 6.1). The VaR defined on these changes features a more stationary evolution than the reserve.

[Insert Figure 6.1: VaR defined from P&L distribution]

The reserve and VaR can equivalently be defined from the **Loss and Profit (L&P)** distribution as:

$$P_t(-W_{t+h} < R) \;=\; 1 - \alpha = \alpha^*, \tag{6.5}$$
$$P_t(W_t - W_{t+h} < VaR) \;=\; 1 - \alpha = \alpha^*. \tag{6.6}$$

From (6.5), the level of reserve is the $(1-\alpha)$ -quantile of the loss and profit distribution. In this approach, $\alpha^* = 1 - \alpha$ takes large values such as 99%, 95%, or 90%. Definitions (6.1)-(6.4) and (6.5)-(6.6) are equivalent, but their choice depends on the interest. The P&L definition is generally considered by banks, which focus on profits, whereas the L&P definitions are typically adopted by regulators, which are more concerned in controlling losses.

In case of primary products such as consumer loans, or mortgages, which are not directly traded on financial markets, the benchmark value W_t corresponds to the accounting value, which is generally the contractual value of the loan or mortgage. This value does not account for default risk and systematically overestimate the "true" value of the portfolio. The difference $W_t - W_{t+h}$ measures the loss due to default and is always positive. The associated VaR is called **CreditVaR** (see Figure 6.2).

[Insert Figure 6.2: CreditVaR defined from L&P distribution]

Example 6.1: Gaussian P&L distribution

Let us consider a Gaussian P&L distribution:

$$W_{t+h}|I_t \sim N(m_t, \sigma_t^2),$$

where I_t is the available information, m_t and σ_t^2 the conditional mean and variance of the future portfolio value, respectively. We have:

$$
\begin{aligned}
& P_t[W_{t+h} + R < 0] = \alpha \\
\iff \quad & P_t[m_t + \sigma_t Z + R < 0] = \alpha \\
\iff \quad & P_t\left(Z < -\frac{R + m_t}{\sigma_t}\right) = \alpha \\
\iff \quad & \Phi\left(-\frac{R + m_t}{\sigma_t}\right) = \alpha \\
\iff \quad & -R = m_t + \sigma_t \Phi^{-1}(\alpha) = Q_t(\alpha), \text{ say,}
\end{aligned} \tag{6.7}
$$

where Z denotes a standard normal variable, and

$$
Q_t(\alpha) = m_t + \sigma_t \Phi^{-1}(\alpha), \tag{6.8}
$$

is the quantile function, that is, the inverse of the cdf of the profit and loss distribution. Since α is small in practice, $\Phi^{-1}(\alpha)$ is negative and, from (6.7), we see that the reserve diminishes when the expected portfolio value increases, and increases when its variance increases. In practice, the P&L distributions often feature fat tails and the Gaussian model above is inappropriate, leading to an underestimation of the risk and of the reserve.

The levels of the reserves, that are the quantiles, are natural measures of risk. They are rather easy to understand by professionals and are largely used in the industry. From the regulatory point of view, the required capital has to be defined without ambiguity, and therefore I_t, h, α have to be fixed. These quantities typically depend on the type of risk, that is, market risk, credit risk, etc., and on the level of sophistication of the risk model. For instance, for credit risk in the "standard approach" of Basel 2 (Pillar 1) one generally chooses $h = 1$ year and $\alpha = 5\%$, while the information set I_t corresponds to the absence of information. Then, the P&L distribution reduces to an unconditional distribution, called **historical** in Basel 2 terminology. However, for internal models of risk management (which correspond to Pillar 2 of Basel 2) and in the "advanced approach", several conditional quantiles have to be followed jointly to take into account the effect of information, the term structure of risk and the more or less severe risk control.

ii) Distortion Risk Measures

The set of all quantiles risk measures, that are the VaR's for all risk levels α and terms h, is highly informative, since it provides the entire P&L distribution for all horizons. However, when a single VaR is selected, for instance to define the required capital, this risk measure has a drawback. Indeed, it accounts for the probability of loss, but not for the magnitude of the loss, when a loss arises. An extended set of risk measures is obtained by considering weighted combinations of opposite quantiles.

Definition 6.1: *Let us consider a profit and loss distribution with quantile function Q, and a positive probability measure H. A distortion risk measure (DRM) is defined by:*

$$\pi(Q, H) = -\int_0^1 Q(u)dH(u).$$

The measure H is called the distortion measure.

This family of risk measures has been extensively studied in a series of papers by Wang [Wang (1996), (2000)]. In order a DRM to have desirable properties as a risk measure, function H has to be concave (see Review Appendix B.3).

Example 6.2: VaR

The Value-at-Risk at level α corresponds to the limiting case of a point mass distortion measure $H_\alpha(u) = \mathbb{1}_{u \geq \alpha}$ (see Figure 6.3). The lack of concavity of the indicator function explains some drawbacks of the VaR (α) used as a single measure of risk.

[Insert Figure 6.3: VaR distortion measure]

Example 6.3: TailVaR

When $H_\alpha(u) = \min(u/\alpha, 1)$, with $\alpha \in (0, 1)$ (see Figure 6.4), we get the equally weighted average of VaR on the interval $[0, \alpha]$:

$$\begin{aligned}
\pi(Q, H_\alpha) &= -\int_0^\alpha \frac{Q(u)}{\alpha} du \\
&= -\frac{1}{\alpha}\int_0^{Q(\alpha)} v dF(v) \quad \text{[by change of variable } v = Q(u)\text{]} \\
&= E[-W | W < Q(\alpha)],
\end{aligned}$$

which measures the expected loss behind the VaR. This measure is also called the **Expected Shortfall** [Acerbi, Tasche (2002)].

[Insert Figure 6.4: Distortion measure for the TailVaR]

6.2 Local Analysis of a Quantile Function

Let us now analyse the sensitivity of a risk measure with respect to a small change in the P&L distribution.

i) Bahadur's expansion

Let us consider a sequence of one-dimensional continuous distributions with cdf F_n and positive density f_n, tending to a probability distribution with cdf F and positive density f, as the index n tends to infinity. We assume that the limiting density is differentiable. The assumptions above imply the existence of the quantile functions Q_n and Q defined by:

$$F_n[Q_n(u)] = u, \quad F[Q(u)] = u, \quad \forall u \in (0, 1). \tag{6.9}$$

Proposition 6.1: (Bahadur's expansion) *We have:*

$$Q_n(u) - Q(u) \simeq -\frac{F_n[Q(u)] - F[Q(u)]}{f[Q(u)]}.$$

Proof: From definitions (6.9) of the quantile functions, we deduce that:

$$\begin{aligned}
0 &= F_n[Q_n(u)] - F[Q(u)] \\
&= F_n[Q_n(u)] - F_n[Q(u)] + F_n[Q(u)] - F[Q(u)],
\end{aligned}$$

and thus:

$$F_n[Q_n(u)] - F_n[Q(u)] = -\left(F_n[Q(u)] - F[Q(u)]\right).$$

By considering a first-order expansion of the left hand side, we get:

$$f_n[Q(u)][Q_n(u) - Q(u)] \simeq -\left(F_n[Q(u)] - F[Q(u)]\right).$$

Equivalently we have:

$$\begin{aligned} Q_n(u) - Q(u) &\simeq -\frac{F_n[Q(u)] - F[Q(u)]}{f_n[Q(u)]} \\ &\simeq -\frac{F_n[Q(u)] - F[Q(u)]}{f[Q(u)]}, \end{aligned}$$

which is the result of the Proposition.

QED

The first-order expansion of the quantile function has been derived in Bahadur (1966) and is largely used in nonparametric estimation of the quantile function [see e.g. Koenker (2005)], or equivalently in nonparametric estimation of the VaR [see e.g. Gouriéroux (2009)]. As seen in Section 6.3, in the applications to risk measures, the difference $F_n - F$ is of order $1/n$, and thus by Proposition 6.1, the same order is expected for the difference between the quantile functions.

ii) Interpretation in terms of variables

It is useful to give an interpretation of Bahadur's expansion in terms of random variables. For this purpose, let us assume that F_n is the distribution of a sum $Y_n = Y + W_n$, where variable W_n tends to zero, when n tends to infinity. Then the limiting distribution F is simply the distribution of Y. The proposition below is proved in Appendix 6.6.

Proposition 6.2: *Let us assume that F_n is the distribution of $Y_n = Y + W_n$, with W_n tending to zero as $n \to \infty$. Let us also assume that Y has* *a continuous distribution with positive differentiable density f, that W_n is second-order integrable conditional on $Y = y$, and that $E(W_n^2|Y = y)$ is differentiable with respect to y. Then, the Bahadur's expansion can also be written as:*

$$\begin{aligned} Q_n(u) - Q(u) \simeq\ & E[W_n|Y = Q(u)] \\ & -\frac{1}{2}\frac{d \log f[Q(u)]}{dy} E[W_n^2|Y = Q(u)] - \frac{1}{2}\frac{\partial E[W_n^2|Y = Q(u)]}{\partial y}. \end{aligned} \tag{6.10}$$

Equivalently, the sum of the last two terms in the right hand side of approximation (6.10) is equal to:

$$-\frac{1}{2f[Q(u)]}\frac{d}{dy}\left\{f(y)E[W_n^2|Y = y]\right\}_{y=Q(u)}.$$

The interpretation in terms of random variables has been first derived in the literature when $W_n = \varepsilon_n W$, where ε_n is a scalar tending to zero [see Gouriéroux, Laurent, Scaillet (2000), Wilde (2001), Martin, Wilde (2003)]. In this case, we get:

$$\begin{aligned} Q_n(u) - Q(u) \simeq\ & \varepsilon_n E[W|Y = Q(u)] \\ & -\frac{\varepsilon_n^2}{2}\left(\frac{\partial \log f[Q(u)]}{\partial y} E[W^2|Y = Q(u)] + \frac{\partial E[W^2|Y = Q(u)]}{\partial y}\right). \end{aligned} \tag{6.11}$$

In the granularity framework, we have $\varepsilon_n = 1/\sqrt{n}$ and $E[W|Y = Q(u)] = 0$. Thus, only the second component of the right hand side of (6.11) matters (see Sections 6.3 and 6.4).

iii) Local analysis of a distortion risk measure

This local analysis is immediately deduced from the definition of a DRM. Indeed, we have:

$$\begin{aligned}\pi(Q_n, H) - \pi(Q, H) &= -\int_0^1 [Q_n(u) - Q(u)]dH(u) \\ &\simeq \int_0^1 \frac{F_n[Q(u)] - F[Q(u)]}{f[Q(u)]} dH(u),\end{aligned}$$

if we consider the expression of Bahadur's expansion in Proposition 6.1.

6.3 Granularity Adjustment in the Static Model

In this section, we consider a large portfolio of homogenous risks $y_{1,t+1}, \ldots, y_{n,t+1}$, satisfying the assumption of exchangeability. Conditionally on the factor f_{t+1}, the risks are i.i.d. with density $h(y_{i,t+1}|f_{t+1})$. The future portfolio value is $W_{t+1} = \sum_{i=1}^{n} y_{i,t+1}$. Let us assume that the underlying factor values f_t, with t varying, are i.i.d. with density g, which corresponds to a static framework. The P&L density of W_{t+1} is:

$$\int h^{*n}(w|f_{t+1})g(f_{t+1})df_{t+1}, \qquad (6.12)$$

where h^{*n} denotes the $n-th$ convoluate[1] of density $h(.|f_{t+1})$. It is difficult to compute the $n-th$ convoluate, which involves a $n-1$ dimensional integral. The aim of this section is to derive an approximation of the P&L distribution valid up to order $1/n$ and to deduce the corresponding approximation of the risk measures by applying the results of Section 6.2.

We first consider the static Gaussian linear factor model. Then the results are extended to the general static framework. In this section, we consider the level of reserve by individual asset, that is, the total reserve divided by n. Equivalently, we focus on the quantile of $W_{t+1}/n = \bar{y}_{n,t+1}$.

[1] That is the density of the sum of independent random variables with identical distribution.

i) Static Gaussian linear factor model

The individual risks are:

$$y_{i,t+1} = F_{t+1} + u_{i,t+1}, \quad i = 1, \ldots, n,$$

where $F_{t+1} \sim N(m, \sigma^2)$, $u_{i,t+1} \sim IIN(0, \eta^2)$, and F_{t+1} and $u_{i,t+1}$ for $i = 1, \ldots, n$ are serially and cross-sectionally independent. In this simple Gaussian framework, the P&L distribution (6.11) is known in closed form:

$$W_{t+1} \sim N[nm, n^2\sigma^2 + n\eta^2],$$

and:

$$W_{t+1}/n \sim N(m, \sigma^2 + \eta^2/n).$$

The quantile function for W_{t+1}/n is (see Example 6.1):

$$Q_n(u) = m + (\sigma^2 + \eta^2/n)^{1/2}\Phi^{-1}(u).$$

Its first-order expansion in $1/n$ is:

$$\begin{aligned} Q_n(u) &= m + \sigma\Phi^{-1}(u) + \frac{1}{2n}\frac{\eta^2}{\sigma}\Phi^{-1}(u) + o(1/n) \\ &= Q(u) + \frac{1}{2n}\frac{\eta^2}{\sigma}\Phi^{-1}(u) + o(1/n). \end{aligned} \tag{6.13}$$

The quantile function $Q(u)$ corresponds to a limit portfolio of infinite size with Gaussian P&L distribution $N(m, \sigma^2)$. Then, the second term in the right hand side of (6.13) gives the GA for the quantile at order $1/n$. This GA depends on the loss probability u and on the ratio of the common risk variance and the idiosyncratic risk volatility.

The GA for the quantile can also be derived by considering the Bahadur's expansion. We have:

$$W_{t+1}/n = F_{t+1} + \frac{1}{n}\sum_{i=1}^{n} u_{i,t+1} = F_{t+1} + \bar{u}_{n,t+1},$$

where $\bar{u}_{n,t+1}|F_{t+1} \sim N(0, \eta^2/n)$. In particular we have:

$$E(\bar{u}_{n,t+1}|F_{t+1}) = 0, \quad E(\bar{u}^2_{n,t+1}|F_{t+1}) = \eta^2/n, \quad \frac{d}{df}E(\bar{u}^2_{n,t+1}|F_{t+1} = f) = 0.$$

By using Proposition 6.2 with $Y = F_{t+1}$ and $W_n = \bar{u}_{n,t+1}$, we get:

$$Q_n(u) \simeq Q(u) - \frac{1}{2}\frac{d \log f[Q(u)]}{dy}\frac{\eta^2}{n},$$

where f and Q are the pdf and the quantile function of F_{t+1}, respectively. Now, we know that:

$$\log f(y) = -\frac{1}{2}\log 2\pi - \frac{1}{2}\log \sigma^2 - \frac{1}{2}\frac{(y-m)^2}{\sigma^2},$$

and $Q(u) = m + \sigma\Phi^{-1}(u)$. We deduce that:

$$\frac{d \log f[Q(u)]}{dy} = -\frac{Q(u) - m}{\sigma^2} = -\frac{\Phi^{-1}(u)}{\sigma}.$$

By substitution, we recover formula (6.13).

ii) The general static framework

Let us now consider the general static framework of iid factor values. The factor is not necessarily Gaussian, can be multivariate, and the relation between factor and individual risks can be nonlinear. The standardized P&L is the distribution of $W_{t+1}/n = \frac{1}{n}\sum_{i=1}^{n} y_{i,t+1}$. Since the individual risks are independent and identically distributed given the future factor value, we have at order $1/n$:

$$(W_{t+1}/n)|F_{t+1} \approx N[m(F_{t+1}), \sigma^2(F_{t+1})/n], \tag{6.14}$$

where

$$m(F_{t+1}) = E[y_{i,t+1}|F_{t+1}] \quad \text{and} \quad \sigma^2(F_{t+1}) = V[y_{i,t+1}|F_{t+1}]. \tag{6.15}$$

The approximation (6.14) is derived by applying the CLT in the cross-section at date $t+1$ **conditional on the factor value** F_{t+1}.

Equivalently, we can write:

$$W_{t+1}/n \simeq m(F_{t+1}) + \frac{\sigma(F_{t+1})}{\sqrt{n}} Z, \tag{6.16}$$

where $Z \sim N(0,1)$ is independent of F_{t+1}. Then, by applying Proposition 6.2 with $Y = m(F_{t+1})$ and $W_n = \frac{\sigma(F_{t+1})}{\sqrt{n}} Z$, we have:

$$E\left(\frac{\sigma(F_{t+1})Z}{\sqrt{n}} | m(F_{t+1})\right) = \frac{1}{\sqrt{n}} E[\sigma(F_{t+1})|m(F_{t+1})]E(Z) = 0,$$

and:

$$\begin{aligned} E\left(\frac{\sigma^2(F_{t+1})Z^2}{n} | m(F_{t+1})\right) &= E\left[\frac{\sigma^2(F_{t+1})Z^2}{n} | m(F_{t+1})\right] \\ &= E\left[\frac{\sigma^2(F_{t+1})}{n} | m(F_{t+1})\right] E(Z^2) \\ &= \frac{1}{n} E[\sigma^2(F_{t+1})|m(F_{t+1})], \end{aligned}$$

where we have used the independence between Z and F_{t+1}. We deduce the GA for the VaR.

Proposition 6.3: *In a static factor model we have:*

$$\begin{aligned} Q_n(u) - Q(u) &\simeq -\frac{1}{2n}\left(\frac{\partial \log f[Q(u)]}{\partial y} E[\sigma^2(F_{t+1})|m(F_{t+1}) = Q(u)] \right. \\ &\quad \left. + \frac{\partial E\left[\sigma^2(F_{t+1})|m(F_{t+1}) = Q(u)\right]}{\partial y}\right), \end{aligned}$$

where f (resp. Q) is the pdf (resp. the quantile function) of $m(F_{t+1})$.

The quantile $Q(u)$ is the CSA approximation. This quantile is computed for the limit (virtual) portfolio of infinite size and corresponds to the quantile of the distribution of $m(F_{t+1})$. Indeed, by the LLN applied in the cross-section at date $t+1$ conditional on F_{t+1}, the standardized portfolio value $W_{n,t+1}$ converges to the conditional mean $m(F_{t+1})$. The CSA approximation of the portfolio quantile by means of function $Q(u)$ corresponds to the Vasicek (1991) approach. The GA $Q_n(u) - Q(u)$ in Proposition 6.3 is the correction at order $1/n$ for the finite portfolio size. This GA involves the density of $m(F_{t+1})$, the conditional mean of volatility $\sigma^2(F_{t+1})$ given $m(F_{t+1})$, as well as their first-order derivative, evaluated at the loss level $Q(u)$.

Example 6.4: Let us consider a portfolio of zero-coupon corporate bonds with time-to-maturity 1 and same nominal equal to 1. The risk variables $y_{i,t+1}$ are dichotomous and correspond to the individual default events. Then, W_{t+1}/n corresponds to the portfolio loss per bond. The risk variables $y_{i,t+1}$ are conditionally i.i.d. with Bernoulli distribution $\mathcal{B}(1, F_{t+1})$ given F_{t+1}. The factor F_{t+1} admits values in $(0, 1)$ and corresponds to the stochastic default probability at date $t+1$. We have $m(F_{t+1}) = F_{t+1}$ and $\sigma^2(F_{t+1}) = F_{t+1}(1 - F_{t+1})$. From Proposition 6.3 we deduce:

$$
\begin{aligned}
& Q_n(u) - Q(u) \\
\simeq\ & -\frac{1}{2n}\left\{\frac{d\log f[Q(u)]}{dy}E[\sigma^2(F_{t+1})|F_{t+1} = Q(u)] + \frac{\partial E[\sigma^2(F_{t+1})|F_{t+1} = Q(u)]}{\partial y}\right\} \\
=\ & -\frac{1}{2n}\left[\frac{d\log f[Q(u)]}{dy}Q(u)[1 - Q(u)] + 1 - 2Q(u)\right],
\end{aligned}
$$

where f (resp. Q) is the density of F_{t+1} (resp. the quantile function). When $\Phi^{-1}(F_t)$ follows a Gaussian distribution $N(\mu, \eta^2)$, we get $f(y) = \frac{1}{\eta}\phi\left[\frac{\Phi^{-1}(y) - \mu}{\eta}\right]/\phi[\Phi^{-1}(y)]$ and $Q(u) = \Phi[\mu + \eta\Phi^{-1}(u)]$, for $y, u \in (0, 1)$. For instance, in the single risk factor model for default based on the Merton (1974) and Vasicek (1991) structural models, we have [see Section 3.1, in particular equation (3.6)]:

$$
\mu = \frac{\Phi^{-1}(PD)}{\sqrt{1-\rho}}, \quad \eta^2 = \frac{\rho}{1-\rho},
$$

where PD is the unconditional default probability and ρ is the asset correlation. Then, we get the CSA quantile function [see Gagliardini, Gourieroux

(2010)]:

$$Q(u) = \Phi\left[\frac{\Phi^{-1}(PD) + \sqrt{\rho}\Phi^{-1}(u)}{\sqrt{1-\rho}}\right], \tag{6.17}$$

and the GA:

$$GA = \frac{1}{2n}\left\{\frac{Q(u)[1-Q(u)]}{\phi(\Phi^{-1}[Q(u)])}\left(\sqrt{\frac{1-\rho}{\rho}}\Phi^{-1}(u) - \Phi^{-1}[Q(u)]\right) + 2Q(u) - 1\right\}. \tag{6.18}$$

iii) A discussion of the GA order

As for estimation (Chapters 2-4) and prediction (Chapter 5), the GA for the quantile is of order $1/n$. This is due to the unobservable factor. To clarify this point, let us consider the Gaussian model without factor, that is $\sigma^2 = 0$ in Section 6.3 i). The distribution of the standardized portfolio value becomes:

$$W_{t+1}/n \sim N(m, \eta^2/n).$$

The cdf of this distribution is $\Phi\left(\frac{y-m}{\eta/\sqrt{n}}\right)$. Its quantile function is $Q_n(u) = m + \frac{\eta}{\sqrt{n}}\Phi^{-1}(u)$, and its Laplace transform is:

$$\psi(u) = E[\exp(uW/n)] = \exp\left(um + \frac{u^2\eta^2}{n}\right).$$

The first terms in the expansions of these functions with respect to $1/n$ have different orders. This order is $1/\sqrt{n}$ for the quantile function, $1/n$ for the Laplace transform. The convergence is very fast and the order depends on argument y for the cdf, whereas we expect a uniform order for deducing from Bahadur's expansion a uniform order for the quantile function.

To understand why the order is $1/n$ uniformly when an unobservable factor is introduced, while the order can be varying without this unobservable factor, let us consider the expectation of a twice continuously differentiable function a of W_{t+1}/n. We have:

$$\begin{aligned}
& E[a(W_{t+1}/n)] \\
= \; & E[a(m + \frac{\eta}{\sqrt{n}}Z)], \text{ where } Z \sim N(0,1), \\
\simeq \; & a(m) + \frac{da(m)}{dm}\frac{\eta}{\sqrt{n}}E(Z) + \frac{1}{2}\frac{d^2a(m)}{dm^2}\frac{\eta^2}{n}E(Z^2) + o(1/n) \\
= \; & a(m) + \frac{\eta^2}{2n}\frac{d^2a(m)}{dm^2} + o(1/n). \qquad (6.19)
\end{aligned}$$

The difficulty encountered for the cdf and quantile function in a model without unobservable factor is due to the interpretation of the cdf as an expectation of an indicator function:

$$F_n(y) = E[\mathbb{1}_{W_{t+1}/n<y}] = E[\mathbb{1}_{m+\frac{\eta}{\sqrt{n}}Z<y}].$$

Since the indicator function is not differentiable, the expansion (6.19) does not apply. Let us now consider the model with factor $F_{t+1} \sim N(0, \sigma^2)$, say. We have:

$$F_n(y) = E[\mathbb{1}_{W_{t+1}/n<y}] = E[\mathbb{1}_{m+F_{t+1}+\frac{\eta}{\sqrt{n}}Z<y}] = E[a(Z)],$$

where:

$$a(Z) \;=\; P\left[m + F_{t+1} + \frac{\eta}{\sqrt{n}}Z < y|Z\right] = \Phi\left(\frac{y - m - \frac{\eta}{\sqrt{n}}Z}{\sigma}\right).$$

This intermediate integration with respect to the factor transforms the discontinuous indicator function into the smooth conditional probability function. This smoothing explains why the GA order $1/n$ is uniform in a factor model.

iv) Illustration: CSA and GA VaR in the static single risk factor model for default

In this subsection we study the CSA and GA approximations in a numerical illustration for the static single risk factor model for corporate default (see Example 6.4 and Section 3.1). Let us first consider given values for the annual unconditional default probability PD and unconditional asset correlation, that are $PD = 0.01$ and $\rho = 0.12$. They correspond to the annual unconditional default probability of a firm with rating about BB, and to the smallest value of asset correlation suggested in Basel 2 [see BCBS (2001), (2003)]. The CSA and GA approximations for the portfolio VaR in equations (6.17) and (6.18) are displayed in Figure 6.5 as functions of the confidence level u, for u close to 1. The GA approximations are for portfolio sizes $n = 25$, 100 and 1000.

[Insert Figure 6.5: CSA and GA approximations in the static single risk factor model]

As expected, the approximated quantiles are increasing with respect to the confidence level. Moreover, the GA quantile curves are above the CSA quantile curve, and the granularity adjustment is decreasing with respect to the portfolio size because of diversification of the unsystematic risk component. For portfolio size $n = 1000$, the GA approximation is very close to the CSA approximation, while for $n = 100$ the granularity adjustment is quite important.

In order to assess the accuracy of the GA approximation, for some confidence levels we display in Figure 6.5 quantiles computed by Monte-Carlo simulation (see Review Appendix A.1). These quantiles are the empirical quantiles for a simulated sample of $500,000$ replications of the portfolio loss. The discrepancy between the GA approximation and the simulated quantiles decreases with the portfolio size n, and is already rather small for portfolio size $n = 100$. For small values of portfolio size such as $n = 25$, the discontinuity of the portfolio quantile w.r.t. the confidence level can be clearly seen from the simulated quantile values.

The GA approximation can be used to study the behaviour of the portfolio risk measure as a function of the model parameters, that are the unconditional probability of default PD and the asset correlation ρ. Such a study would be very time consuming if performed using Monte-Carlo simulation. In Figure 6.6 we display the CSA approximation and the GA as functions of ρ, for different values of PD. In Figure 6.7 we display the CSA approximation and the GA as functions of PD, for different values of ρ. The portfolio size is $n = 1000$ and the confidence level is $1 - \alpha = 0.99$.

[Insert Figure 6.6: CSA VaR and GA in the static single risk factor model as functions of the asset correlation]

[Insert Figure 6.7: CSA VaR and GA in the static single risk factor model as functions of the probability of default]

Figure 6.6 shows that the CSA VaR is monotone increasing w.r.t. asset correlation ρ, when the probability of default is such that $PD \geq \alpha$; for $PD < \alpha$, the CSA VaR is not monotone w.r.t. ρ and it converges to zero as ρ approaches 1. The granularity adjustment is decreasing w.r.t. asset correlation ρ, when ρ is not close to 1. Figure 6.7 shows that the CSA VaR is monotone increasing w.r.t. the probability of default PD. The granularity adjustment features instead an inverse-U shape. The maximum GA occurs for values of PD corresponding to speculative grade ratings, when ρ is between 0.12 and 0.24.

6.4 Granularity Adjustment in the Dynamic Model

Let us now consider the dynamic framework, with a factor transition density given by $g(f_t|f_{t-1})$. In this extended framework, two granularity adjustments are required. The first one concerns the theoretical risk measure itself and is the analogue of the adjustment derived in the static framework. The second granularity adjustment is a consequence of the unobservability of the factor values and is derived by using the approximate filtering formula of Chapter 5.

i) Granularity adjustment for the distribution of the portfolio value

We have:

$$W_{t+1}/n = m(F_{t+1}) + \frac{\sigma(F_{t+1})}{\sqrt{n}} Z + O(1/n), \tag{6.20}$$

where functions $m(F_{t+1})$ and $\sigma(F_{t+1})$ are defined as in (6.15), and $Z \sim N(0,1)$ is independent of the factor path. Term $O(1/n)$ is conditionally zero-mean since the normalized portfolio value is an unbiased estimator of $m(F_{t+1})$, conditionally on F_{t+1}. Let us denote:

$$\begin{aligned} a(y; f_t, \varepsilon) &= P[m(F_{t+1}) + \sigma(F_{t+1})\varepsilon < y | F_t = f_t] \\ &= \int \mathbb{1}_{m(f_{t+1})+\sigma(f_{t+1})\varepsilon<y} \; g(f_{t+1}|f_t) df_{t+1}. \end{aligned}$$

The cdf of the standardized portfolio value given the observable information $I_t^* = (\underline{y_{1,t}}, \cdots, \underline{y_{n,t}})$ only is:

$$\begin{aligned} F_n(y) &= P[W_{t+1}/n < y | I_t^*] \\ &= E(P[W_{t+1}/n < y | F_t, Z] | I_t^*) \\ &= E\left[a(y, F_t, \frac{Z}{\sqrt{n}}) | I_t^* \right] + o(1/n). \end{aligned}$$

Then, by applying the GA for the filtering distribution of F_t given in Corollary 5.3, we get:

$$F_n(y) = E\left[a(y, \hat{f}_{nt} + \frac{1}{n}\mu_{nt} + \frac{1}{\sqrt{n}} I_{nt}^{-1/2} Z^*, \frac{Z}{\sqrt{n}}) | I_t^* \right] + o(1/n),$$

where Z^* is a standard normal variable, $\hat{f}_{n,t}$ is the cross-sectional approximation of the factor value, $\mu_{nt} = I_{nt}^{-1} \frac{\partial \log g}{\partial f_t}(\hat{f}_{nt}|\hat{f}_{n,t-1}) + \frac{1}{2} I_{nt}^{-2} K_{nt}$ is the mean adjustment in the filtering distribution and the term $I_{nt}^{-1/2} Z^*$ is the adjustment for the variance. Moreover, the Gaussian variables Z and Z^* are independent, since the first one is due to the Central Limit Theorem, whereas the second one corresponds to the numerical approximation of the filtering distribution which involves no stochastic argument. The expression of the cdf can be expanded at order $1/n$. Since $E[Z]=0$, $E[Z^*]=0$, $E[ZZ^*]=0$, $E[Z^2]=E[(Z^*)^2]=1$, we get:

$$\begin{aligned} F_n(y) &= a(y,\hat{f}_{nt},0)+\frac{1}{n}\frac{\partial a(y,\hat{f}_{nt},0)}{\partial f_t}\mu_{nt} \\ &+ \frac{1}{2n}\left[I_{nt}^{-1}\frac{\partial^2 a(y,\hat{f}_{nt},0)}{\partial f_t^2}+\frac{\partial^2 a[y,\hat{f}_{nt},0)}{\partial \varepsilon^2}\right]+o(1/n). \quad (6.21) \end{aligned}$$

The CSA approximation of the cdf is $a(y,\hat{f}_{nt},0)$. The GA is the sum of two components corresponding to

i) the granularity adjustment for filtering, that is,

$$\frac{\partial a(y,\hat{f}_{nt},0)}{\partial f_t}\mu_{nt}+\frac{1}{2}I_{nt}^{-1}\frac{\partial^2 a(y,\hat{f}_{nt},0)}{\partial f_t^2};$$

ii) the granularity adjustment for the theoretical cdf, that is,

$$\frac{1}{2}\frac{\partial^2 a(y,\hat{f}_{nt},0)}{\partial \varepsilon^2}.$$

Due to the independence between Z and Z^*, there is no cross GA.

ii) Granularity adjustment for the Value-at-Risk

The GA for the VaR is directly deduced from the GA of the cdf by applying the Bahadur's expansion (see Proposition 6.1). Let us consider the conditional mean $m(F_{t+1})$. Its cdf conditional to $F_t=f_t$ is $a(y,f_t,0)$, which is the leading term in the expansion (6.21). The associated quantile function (resp. density function) is denoted by $Q(u;f_t)$ [resp. $f(y;f_t)$]. We get:

$$Q_n(u)\simeq Q(u;\hat{f}_{n,t})+GA_{risk}+GA_{filter},$$

and the GA for the VaR at risk level u is the sum of two components:

i) The granularity adjustment for filtering is

$$GA_{filter} = -\frac{1}{n}\frac{1}{f[Q(u;\hat{f}_{nt})]}\left[\frac{\partial a[Q(u;\hat{f}_{nt}),\hat{f}_{nt},0]}{\partial f_t}\mu_{nt} + \frac{1}{2}I_{nt}^{-1}\frac{\partial^2 a[Q(u,\hat{f}_{nt}),\hat{f}_{n,t},0)}{\partial f_t^2}\right].$$

ii) The granularity adjustment for the theoretical risk measure is:

$$GA_{risk} = -\frac{1}{2n}\frac{1}{f[Q(u,\hat{f}_{nt})]}\frac{\partial^2 a[Q(u,\hat{f}_{nt}),\hat{f}_{nt},0)}{\partial \varepsilon^2}.$$

This latter GA can also be written as [see Gagliardini, Gourieroux (2010)]:

$$\begin{aligned} GA_{risk} &= -\frac{1}{2n}\left\{\frac{\partial \log f[y;\hat{f}_{nt}]}{\partial y}E[\sigma^2(F_{t+1})|m(F_{t+1})=y,F_t=\hat{f}_{n,t}]\right. \\ &\left.\quad + \frac{\partial}{\partial y}E[\sigma^2(F_{t+1})|m(F_{t+1})=y,F_t=\hat{f}_{n,t}]\right\}_{y=Q(u;\hat{f}_{nt})}. \end{aligned}$$

This expression is the analogue of the GA in the static factor model of Section 6.3. The distribution of F_{t+1} is now conditional on the current factor value F_t, and this unobservable value is finally replaced by the cross-sectional approximation $\hat{f}_{n,t}$.

It is interesting to note that the major part of the existing literature has proposed the GA_{risk} component as the total adjustment to be applied to the VaR. In a dynamic model, the computations above show that the other component due to the factor unobservability has the same magnitude and has also to be taken into account as shown in the illustration below.

iii) Illustration to dynamic model with stochastic default and recovery

In this illustration we consider an extension of the single risk factor model for corporate default presented in Section 6.3 iv) to account for the dynamics of the systematic factor and a non-zero recovery rate [see Gagliardini, Gourieroux, Monfort (2010)]. The percentage loss on the loan to firm i at the maturity date $t+1$ is:

$$y_{i,t+1} = \mathbb{1}_{A_{i,t+1}<L_{i,t+1}} \left(1 - \frac{A_{i,t+1}}{L_{i,t+1}}\right) = \left(1 - \frac{A_{i,t+1}}{L_{i,t+1}}\right)^+, \tag{6.22}$$

where $A_{i,t+1}$ and $L_{i,t+1}$ are the stochastic asset value and liability of the firm, and $x^+ = \max\{x, 0\}$ denotes the positive part of x. The loss variable $y_{i,t+1}$ is the product of the default indicator $\mathbb{1}_{A_{i,t+1}<L_{i,t+1}}$, that is equal to 1, when the asset value is below the liability, and 0, otherwise, and the percentage loss given default (LGD), that is, $1 - \frac{A_{i,t+1}}{L_{i,t+1}}$. The dynamics of the log asset/liability ratios of the firms follow a linear single risk factor model:

$$\log\left(\frac{A_{i,t}}{L_{i,t}}\right) = F_t + \sigma u_{i,t}, \tag{6.23}$$

where the idiosyncratic shocks $(u_{i,t})$ are $IIN(0,1)$ across time and firms. The single systematic factor F_t follows an autoregressive Gaussian process:

$$F_t = \mu + \gamma(F_{t-1} - \mu) + \eta\sqrt{1-\gamma^2}\varepsilon_t, \tag{6.24}$$

where the innovations $\varepsilon_t \sim IIN(0,1)$ are independent of $(u_{i,t})$. The model parameters are the volatility of the idiosyncratic shocks $\sigma > 0$, the unconditional mean μ and volatility $\eta > 0$ of the systematic factor, and its autocorrelation coefficient γ. The latter is assumed such that $|\gamma| < 1$ to ensure stationarity. When $\gamma \neq 0$, the systematic risk factor features serial dependence.

As for the static model for corporate default (see Remark 3.1 in Section 3.1), there exists alternative parameterizations of model (6.22)-(6.24) admitting a more direct financial interpretation. More precisely, let us consider the unconditional probability of default:

$$PD = P\left[\log\left(A_{i,t}/L_{i,t}\right) < 0\right] = \Phi\left(-\frac{\mu}{\sqrt{\eta^2+\sigma^2}}\right), \tag{6.25}$$

and the unconditional asset correlation:

$$\rho = corr\left[\log\left(A_{i,t}/L_{i,t}\right), \log\left(A_{j,t}/L_{j,t}\right)\right] = \frac{\eta^2}{\eta^2+\sigma^2}, \tag{6.26}$$

for $i \neq j$, respectively. Furthermore, let us introduce the unconditional expected (percentage) loss given default (ELGD):

$$ELGD = E\left[1 - \frac{A_{i,t}}{L_{i,t}} | \frac{A_{i,t}}{L_{i,t}} < 1\right]. \tag{6.27}$$

Gagliardini, Gourieroux and Monfort (2010) derive an expression for $ELGD$ in terms of the structural parameters σ, μ and η [see also Geske (1977)]. Then, the probability of default PD, the asset correlation ρ, the expected loss given default $ELGD$ and the factor autocorrelation γ provide an equivalent parameterization of the model. In Table 6.1 we display the values of the structural parameters σ, μ and η corresponding to some choices of the reduced form parameters PD, ρ and $ELGD$.

[Insert Table 6.1: Reduced form and structural parameters]

In particular, the values 0.45 and 0.75 of $ELGD$ in Table 6.1 are the values of expected loss given default suggested by Basel 2 regulation [see BCBS (2001), (2003)] for senior debt classes on corporate, sovereigns and banks not secured, and subordinated classes on corporate, sovereigns and banks, respectively.

The CSA and GA quantile approximations are derived from the general results in Section 6.4 ii). We present here some steps of the analysis and invite the reader to refer to Gagliardini, Gourieroux, Monfort (2010) for the detailed derivation.

a) The cross-sectional factor approximation at date t is:

$$\hat{f}_{n,t} = \arg\max_{f_t} \left\{ -\frac{1}{2\sigma^2} \sum_{i: y_{i,t} > 0} [\log(1 - y_{i,t}) - f_t]^2 + (n - n_t) \log \Phi(f_t/\sigma) \right\}, \tag{6.28}$$

where $n_t = \sum_{i=1}^{n} \mathbb{1}_{y_{i,t} > 0}$ denotes the number of defaults at date t. The factor approximation corresponds to the Maximum Likelihood estimator of the mean parameter in a Gaussian Tobit regression model with endogenous variable

$\log(1-y_{i,t})$, mean f_t and variance σ^2. The Gaussian approximation at order $1/n$ of the filtering distribution of the unobservable factor f_t in Corollary 5.3 involves statistics $\hat{f}_{n,t}$, $\hat{f}_{n,t-1}$ and n_t.

b) Functions $m(f_{t+1})$ and $\sigma(f_{t+1})$ can be derived from the Black-Scholes pricing formula by exploiting the put option structure of the loss variable $(1-A_{i,t+1}/L_{i,t+1})^+$ and the conditional log-normality of $A_{i,t+1}/L_{i,t+1}$ given $F_{t+1}=f_{t+1}$. We get:

$$m(f_{t+1}) = \Phi(-f_{t+1}/\sigma) - \exp\left(f_{t+1}+\frac{\sigma^2}{2}\right)\Phi\left(-f_{t+1}/\sigma-\sigma\right), \quad (6.29)$$

and:

$$\begin{aligned}\sigma^2(f_{t+1}) &= m(f_{t+1})[1-m(f_{t+1})] - \exp\left(f_{t+1}+\frac{\sigma^2}{2}\right)\Phi\left(-f_{t+1}/\sigma-\sigma\right)\\ &\quad + \exp(2f_{t+1}+2\sigma^2)\Phi\left(-f_{t+1}/\sigma-2\sigma\right). \qquad (6.30)\end{aligned}$$

Function m is monotone decreasing, since the loss $y_{i,t+1}$ is decreasing w.r.t. the factor value F_{t+1}.

c) Finally, function $a(w,\hat{f}_{n,t},0)$ is given by:

$$\begin{aligned}a(w,\hat{f}_{n,t},0) &= P[m(F_{t+1})\leq w|F_t=\hat{f}_{n,t}] = P[F_{t+1}\geq m^{-1}(w)|F_t=\hat{f}_{n,t}]\\ &= \Phi\left(-\frac{m^{-1}(w)-\mu-\gamma(\hat{f}_{n,t}-\mu)}{\eta\sqrt{1-\gamma^2}}\right), \qquad (6.31)\end{aligned}$$

where m^{-1} denotes the inverse of function m.

In Figure 6.8 we display the CSA and GA VaR approximations, and the GA risk and filtering components, as functions of the cross-sectional approximation of the current factor value.

[Insert Figure 6.8: CSA and GA VaR as a function of the cross-sectional factor approximation]

The parameters are such that the annual default probability is $PD = 0.05$, the asset correlation is $\rho = 0.12$, the expected loss given default is $ELGD =$

0.45 and the factor autocorrelation is $\gamma = 0.5$. The corresponding unconditional mean of the factor is $\mu = 3.05$. The information set is such that $n_t/n = PD$ and $\hat{f}_{n,t} = \mu$, while the confidence level is $1 - \alpha = 0.995$. The CSA VaR is decreasing w.r.t. the factor approximation, since the systematic factor has a positive impact on the asset/liability ratios of the firms. The granularity adjustment is quite small for portfolio size $n = 1000$, but is relevant for portfolio size $n = 100$. By comparing the patterns of the GA risk and filtering components, it is seen that the granularity adjustment comes mostly from filtering, at least when the factor approximation is above the factor mean. Indeed, the filtering GA component accounts for the uncertainty of the cross-sectional factor approximation. When this approximation is above the factor mean, the filtering GA component yields an upward correction of the CSA VaR, which reflects a less optimistic belief on the unobservable factor value compared to the cross-sectional approximation.

In Figure 6.9 we display simulated paths of the default frequency n_t/n, the percentage portfolio loss $W_{n,t}/n$, the factor value f_t and its cross-sectional approximation $\hat{f}_{n,t}$. In Figure 6.10 we display the corresponding simulated paths of the CSA and GA VaR, and of the GA risk and filtering components.

[Insert Figure 6.9: Time series of simulated default frequency, portfolio loss, systematic factor and cross-sectional approximation of the factor]

[Insert Figure 6.10: Time series of simulated CSA VaR, GA VaR, and GA risk and filtering components]

The portfolio size is $n = 100$, the confidence level is $1 - \alpha = 0.995$ and the model parameters are as above. The time series of default frequency and portfolio loss have a similar pattern, since they are driven by the same systematic factor. Moreover, at each date the portfolio loss is smaller than the default frequency because of the non-zero recovery rate. The cross-sectional factor approximation is rather accurate. Figure 6.10 shows that the GA VaR is larger and features a smoother time evolution than the CSA VaR. Moreover, whereas the risk component of the granularity adjustment is always positive and rather stable in time, its filtering component varies quite a lot in time and can eventually take negative values. As already remarked in Figure 6.8, the filtering GA component is responsible for most of the granularity adjustment.

Finally, Gagliardini, Gourieroux and Monfort (2010) perform a backtesting analysis to compare the GA VaR and CSA VaR in terms of the frequency and dynamic pattern of violations, that are, the exceedancies of the realized portfolio loss above VaR. They show that the GA VaR is a more accurate approximation of the true portfolio quantile than the CSA VaR.

6.5 Portfolio of Derivatives Written on a Large Portfolio

The VaR and its granularity adjustments have also to be computed for portfolio of derivatives written on a given large portfolio of individual contracts. These derivatives are called Collateralized Debt Obligations (CDO). Typically, the support of such credit derivatives is a given pool of credits, or of Credit Default Swaps (CDS). Then the derivative payoffs are defined by tranching the (normalized) portfolio value $\bar{W}_{n,t+1} = W_{n,t+1}/n$, that is, by considering payoffs of the type:

$$b_j(\bar{W}_{n,t+1}) = \begin{cases} \bar{W}_{n,t+1}, \quad \text{if } \bar{W}_{n,t+1} \in (a_j, a_{j+1}), \text{ say,} \\ 0, \text{ otherwise,} \end{cases}$$

where the a_j and a_{j+1} are called attachment and detachment points, respectively, or by considering straddles defined by combining appropriately European calls with payoffs:

$$b_j(\bar{W}_{n,t+1}) = (\bar{W}_{n,t+1} - a_j)^+, \quad \text{say.}$$

To get a flavour of the GA for a portfolio of such derivatives, we consider a portfolio of CDO's with maturity $t+1$. Their value at $t+1$ is equal to the payoff, and the future derivative portfolio value is: [2]

$$W^D_{n,t+1} = \sum_{j=1}^{J} b_j(\bar{W}_{n,t+1}),$$

where J denotes the number of CDO's in the portfolio. Thus, from (6.20) we deduce the expansion of the derivative portfolio value as:

$$\begin{aligned} W^D_{n,t+1} &= \sum_{j=1}^{J} b_j[m(F_{t+1}) + \sigma(F_{t+1})\frac{Z}{\sqrt{n}} + O(1/n)] \\ &= \sum_{j=1}^{J} b_j[m(F_{t+1})] + \sum_{j=1}^{J} \frac{db_j}{dm}[m(F_{t+1})]\sigma(F_{t+1})\frac{Z}{\sqrt{n}} \\ &+ \frac{1}{2}\sum_{j=1}^{J} \frac{d^2 b_j}{dm^2}[m(F_{t+1})]\sigma^2(F_{t+1})\frac{Z^2}{n} + O(1/n), \end{aligned}$$

where the $O(1/n)$ term is zero-mean conditional on the factor.

Then, we can apply the GA formula of Proposition 6.2, with $Y = \sum_{j=1}^{J} b_j[m(F_{t+1})]$ as the limiting future derivative portfolio value for infinite n, and the sum W_n, say, of the two other components of the right hand side to capture the next terms in the expansion. By using the moments $E[Z] = 0$, $E[Z^2] = 1$, we get:

$$E(W_n|Y = y) = \frac{1}{2n}E\left[\sum_{j=1}^{J} \frac{d^2 b_j[m(F_{t+1})]}{dm^2}\sigma^2(F_{t+1})|\sum_{j=1}^{J} b_j[m(F_{t+1})] = y\right] + o(1/n),$$

$$E(W_n^2|Y = y) = \frac{1}{n}E\left[\left(\sum_{j=1}^{J} \frac{db_j[m(F_{t+1})]}{dm}\right)^2 \sigma^2(F_{t+1})|\sum_{j=1}^{J} b_j[m(F_{t+1})] = y\right] + o(1/n).$$

[2]When the maturity of the derivatives is strictly larger than 1, the derivatives maturity does not correspond to the selected horizon for the VaR and it is necessary to compute the future derivative price. This can be done by the approximate derivative pricing approach introduced in Chapter 5, accounting in particular for the GA of the derivative prices.

Then, the GA is derived from Proposition 6.2.

6.6 Summary

The recent regulations require the computation of reserves for large portfolios, possibly including derivatives. The required capital is based on risk measures such as the VaR or the Expected Shortfall. Granularity theory is used to derive closed form expressions for the reserves at order $1/n$. The explicit formulas can be introduced in the software for risk management and risk control.

6.7 Appendix: Interpretation of the Bahadur's Expansion

Let us prove Proposition 6.2 when Y and W_n admit a joint pdf [see Gagliardini, Gourieroux (2010) for the general case]. We have:

$$\begin{aligned} -\frac{F_n(y)-F(y)}{f(y)} &= \frac{P(Y<y)-P(Y+W_n<y)}{f(y)} \\ &= \frac{1}{f(y)}\int[\int_{y-w}^{y} f_n(z,w)dz]dw, \end{aligned}$$

where $f_n(y,w)$ denotes the joint density of (Y, W_n). Thus, we deduce:

$$\begin{aligned} &- \frac{F_n(y)-F(y)}{f(y)} \\ &= \frac{1}{f(y)}\int[\int_{y-w}^{y} f_n(y,w)dz]dw + \frac{1}{f(y)}\int[\int_{y-w}^{y}[f_n(z,w)-f_n(y,w)]dz]dw \\ &\simeq \frac{1}{f(y)}\int w f_n(y,w)dw + \frac{1}{f(y)}\int \frac{\partial f_n(y,w)}{\partial y}[\int_{y-w}^{y}(z-y)dz]dw \\ &= E[W_n|Y=y] - \frac{1}{2}E[W_n^2\frac{\partial \log f_n(y,W_n)}{\partial y}|Y=y]. \end{aligned} \tag{6.32}$$

Let us now decompose the joint density into the unconditional density of Y and the conditional density of W_n given Y, i.e. $f_n(y,w) = f(y)f_n(w|y)$, say. We get:

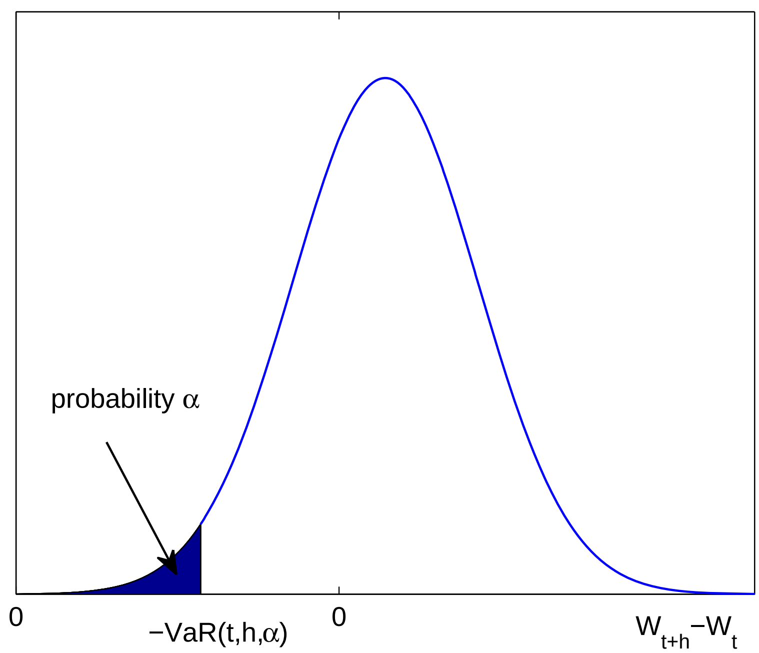
$$\begin{aligned} &- \frac{F_n(y)-F(y)}{f(y)} \\ &\simeq E[W_n|Y=y] - \frac{1}{2}\frac{\partial \log f(y)}{dy}E[W_n^2|Y=y] - \frac{1}{2}\int w^2\frac{\partial \log f_n(w|y)}{\partial y}f_n(w|y)dw \\ &= E[W_n|Y=y) - \frac{1}{2}\frac{\partial \log f(y)}{dy}E(W_n^2|Y=y) - \frac{1}{2}\frac{\partial}{\partial y}(\int w^2 f_n(w|y)dw) \\ &= E(W_n|Y=y) - \frac{1}{2}\frac{\partial \log f(y)}{dy}E(W_n^2|Y=y) - \frac{1}{2}\frac{\partial}{\partial y}E(W_n^2|Y=y). \end{aligned}$$

From Proposition 6.1, the result follows.

Table 6.1: Reduced form and structural parameters.

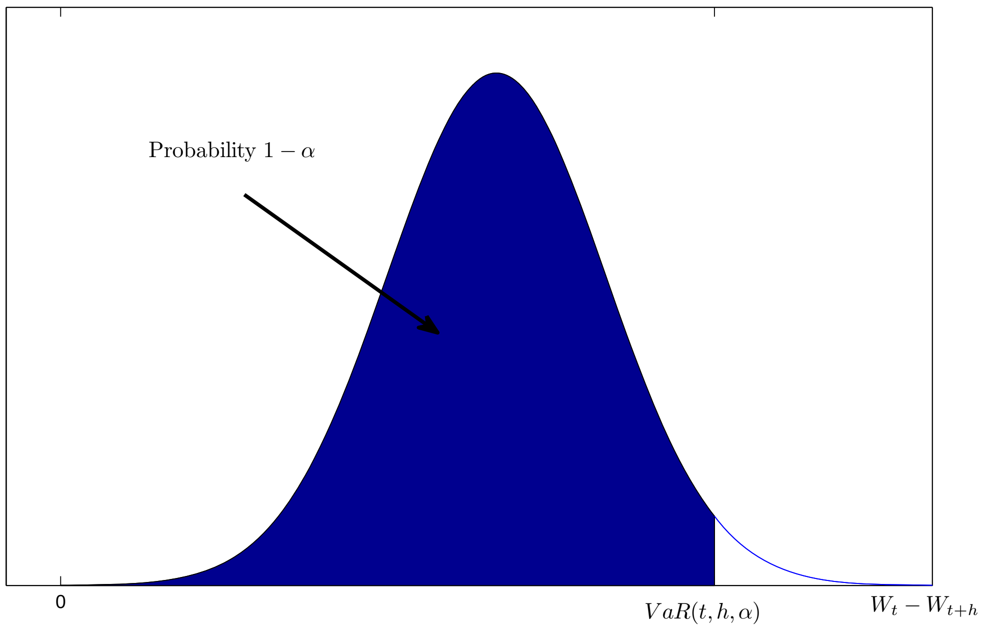
Reduced form parameters			Structural parameters		
$ELGD$	PD	ρ	μ	η	σ
0.45	1.5%	0.12	4.799	0.766	2.074
		0.24	4.799	1.083	1.928
		0.50	4.799	1.564	1.564
0.45	5%	0.12	3.050	0.642	1.739
		0.24	3.050	0.908	1.616
		0.50	3.050	1.311	1.311
0.75	1.5%	0.12	16.993	2.713	7.346
		0.24	16.993	3.836	6.827
		0.50	16.993	5.537	5.537
0.75	5%	0.12	10.669	2.247	6.085
		0.24	10.669	3.178	5.655
		0.50	10.669	4.587	4.587

Figure 6.1: VaR defined from the P&L Distribution.



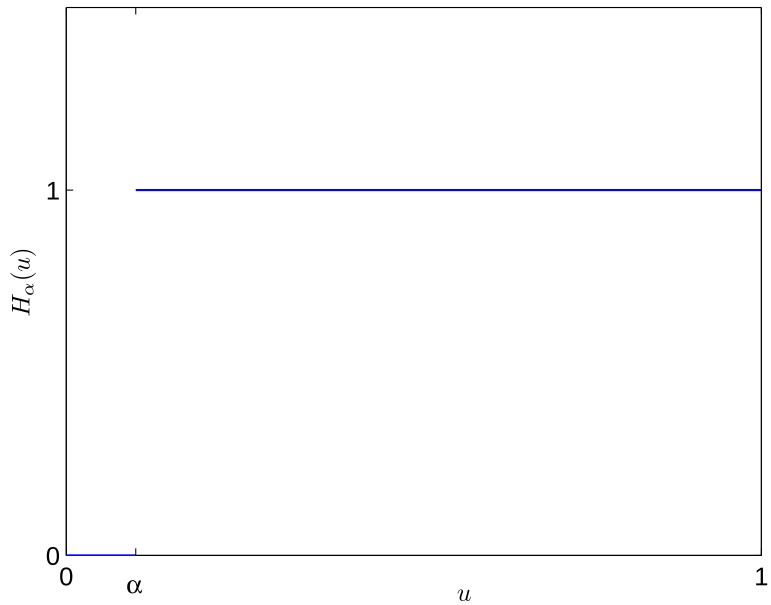
The Value-at-Risk $VaR(t, h, \alpha)$ is the opposite of the quantile at level α of the conditional distribution of $W_{t+h} - W_t$ given date t information (P&L distribution). The shaded area corresponds to a probability of α.

Figure 6.2: CreditVaR defined from the L&P Distribution.



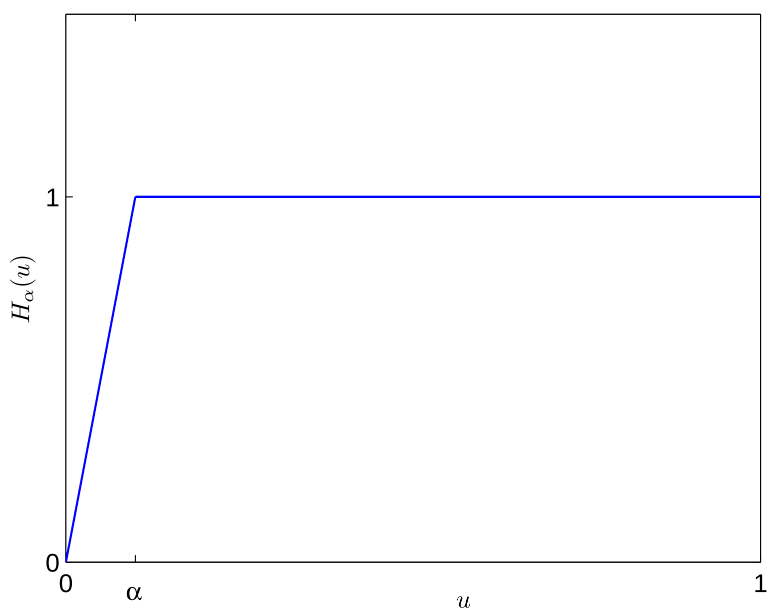
The CreditVaR $VaR(t,h,\alpha)$ is the quantile at level $1-\alpha$ of the conditional distribution of $W_t - W_{t+h}$ given date t information (L&P distribution). The shaded area corresponds to a probability of $1-\alpha$.

Figure 6.3: VaR distortion measure.



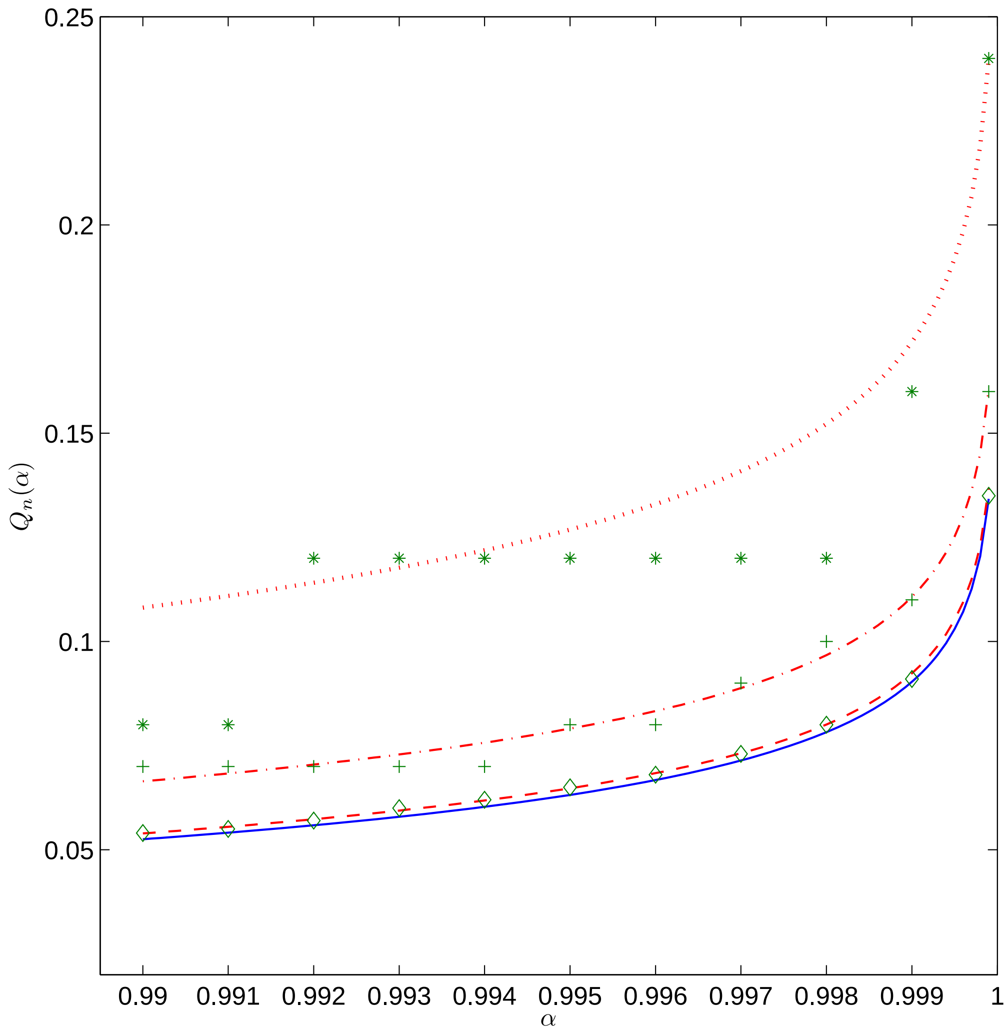
The distortion measure for the VaR at level α is the point mass measure with cdf $H_\alpha(u) = \mathbb{1}_{u\geq\alpha}$.

Figure 6.4: Distortion measure for the TailVaR.



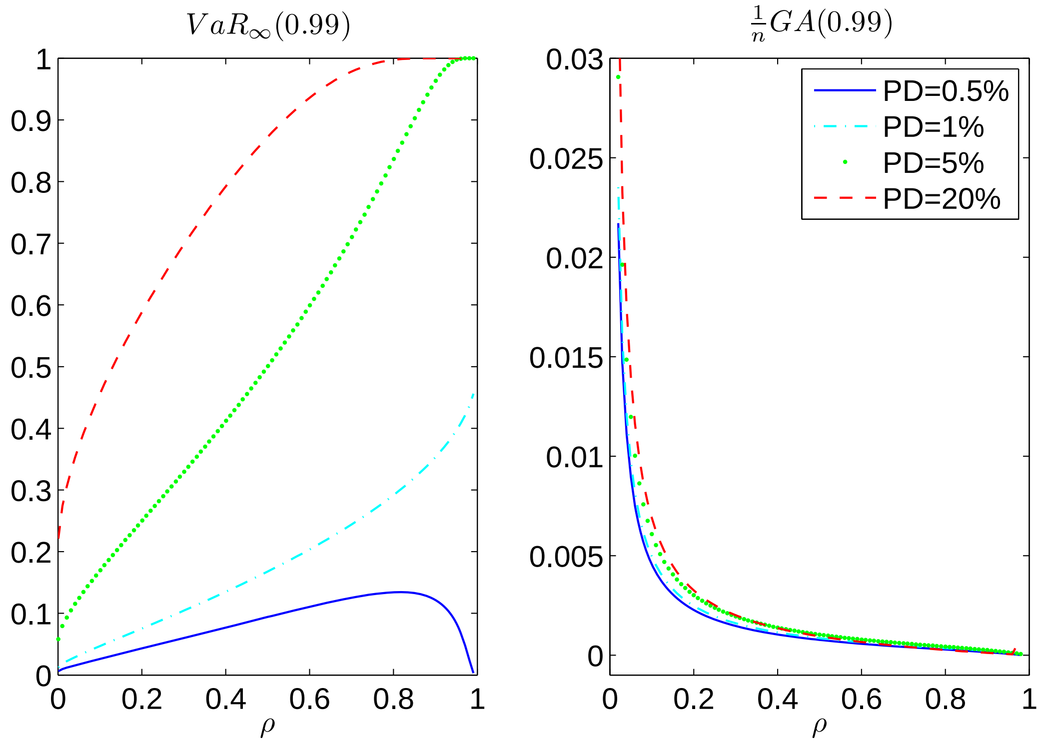
The distortion measure for the TailVaR at level α is the uniform distribution on $[0, \alpha]$ with cdf $H_\alpha(u) = min\{u/\alpha, 1\}$.

Figure 6.5: CSA and GA VaR in the static single risk factor model.



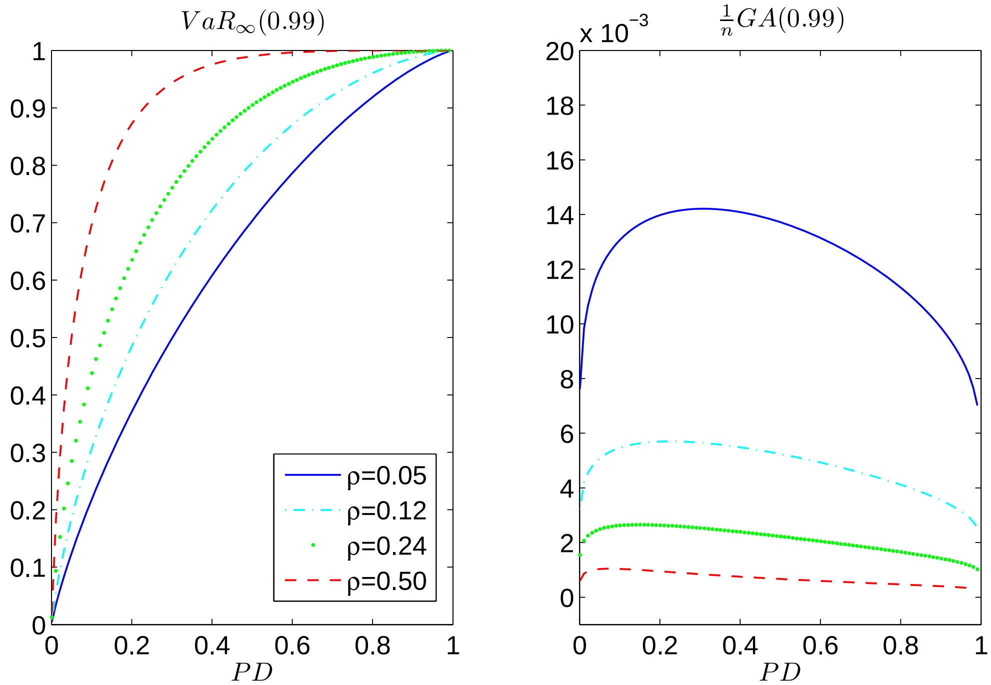
The Figure displays the CSA approximation (solid line) and the GA approximations for the portfolio VaR in a static single risk factor model, as functions of the confidence level α. The GA approximations are for portfolio sizes $n = 25$ (dotted line), $n = 100$ (dashed-dotted line) and $n = 1000$ (dashed line). Stars, crosses and diamonds correspond to quantiles computed with Monte-Carlo simulation based on $500,000$ replications of the portfolio loss, for portfolio sizes $n = 25$, $n = 100$ and $n = 1000$, respectively. The unconditional default probability is $PD = 0.01$ and the asset correlation is $\rho = 0.12$.

Figure 6.6: CSA VaR and GA in the static single risk factor model as functions of asset correlation.



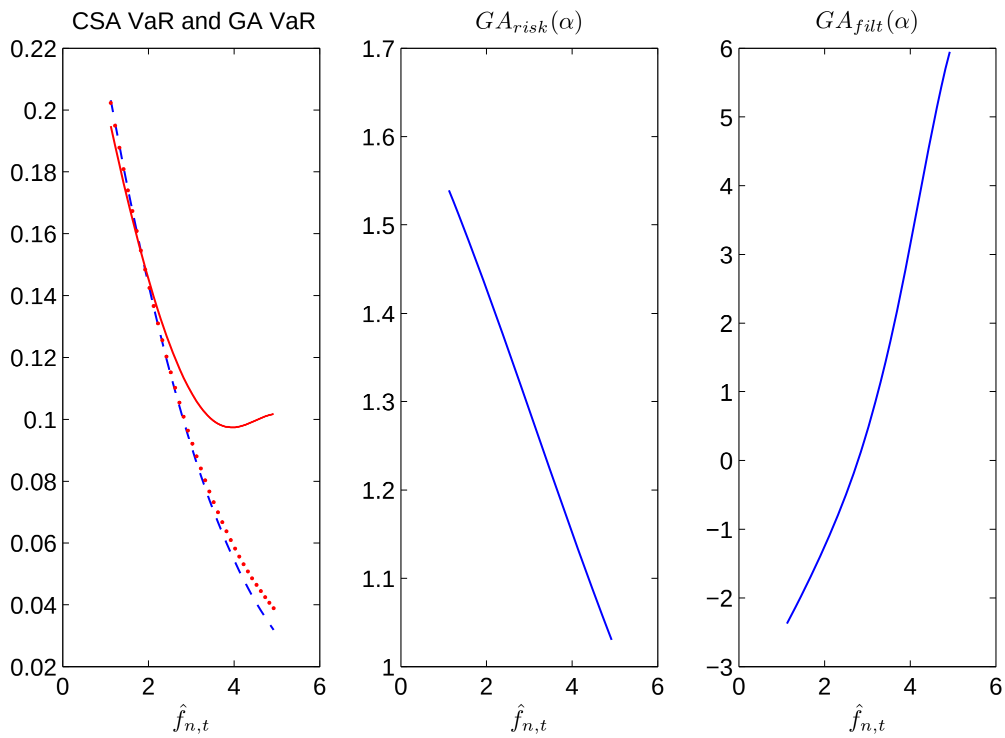
The left panel displays the CSA VaR, and the right panel displays the GA, as functions of asset correlation ρ, for different values of the unconditional default probability, that are $PD = 0.005$, 0.01, 0.05 and 0.20, in the static single risk factor model for default. The portfolio size is $n = 1000$ and the confidence level is $1 - \alpha = 0.99$.

Figure 6.7: CSA VaR and GA in the static single risk factor model as functions of default probability.



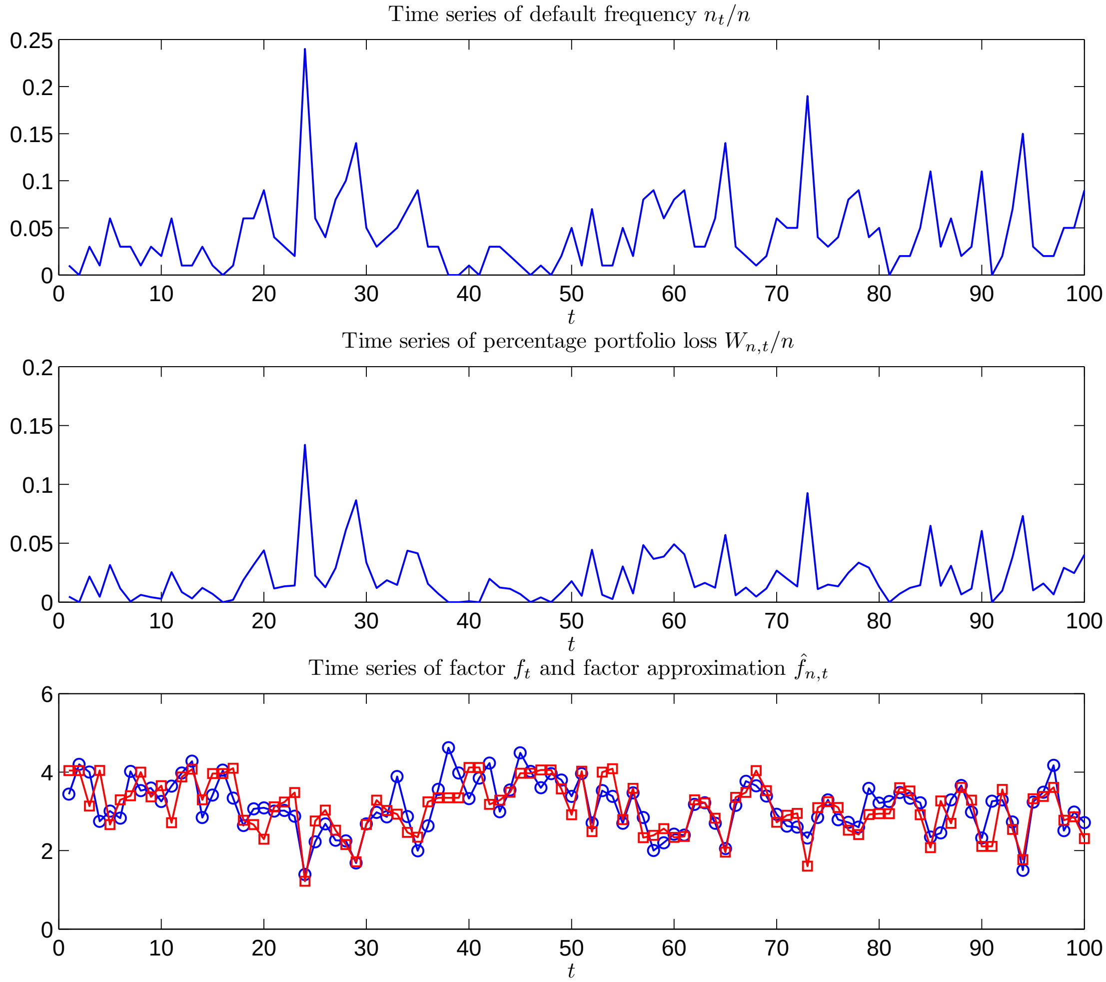
The left panel displays the CSA VaR, and the right panel displays the GA, as functions of the probability of default PD, for different values of the asset correlation, that are $\rho = 0.05$, 0.12, 0.24 and 0.50, in the static single risk factor model for default. The portfolio size is $n = 1000$ and the confidence level is $1 - \alpha = 0.99$.

Figure 6.8: CSA and GA VaR as a function of the cross-sectional factor approximation.



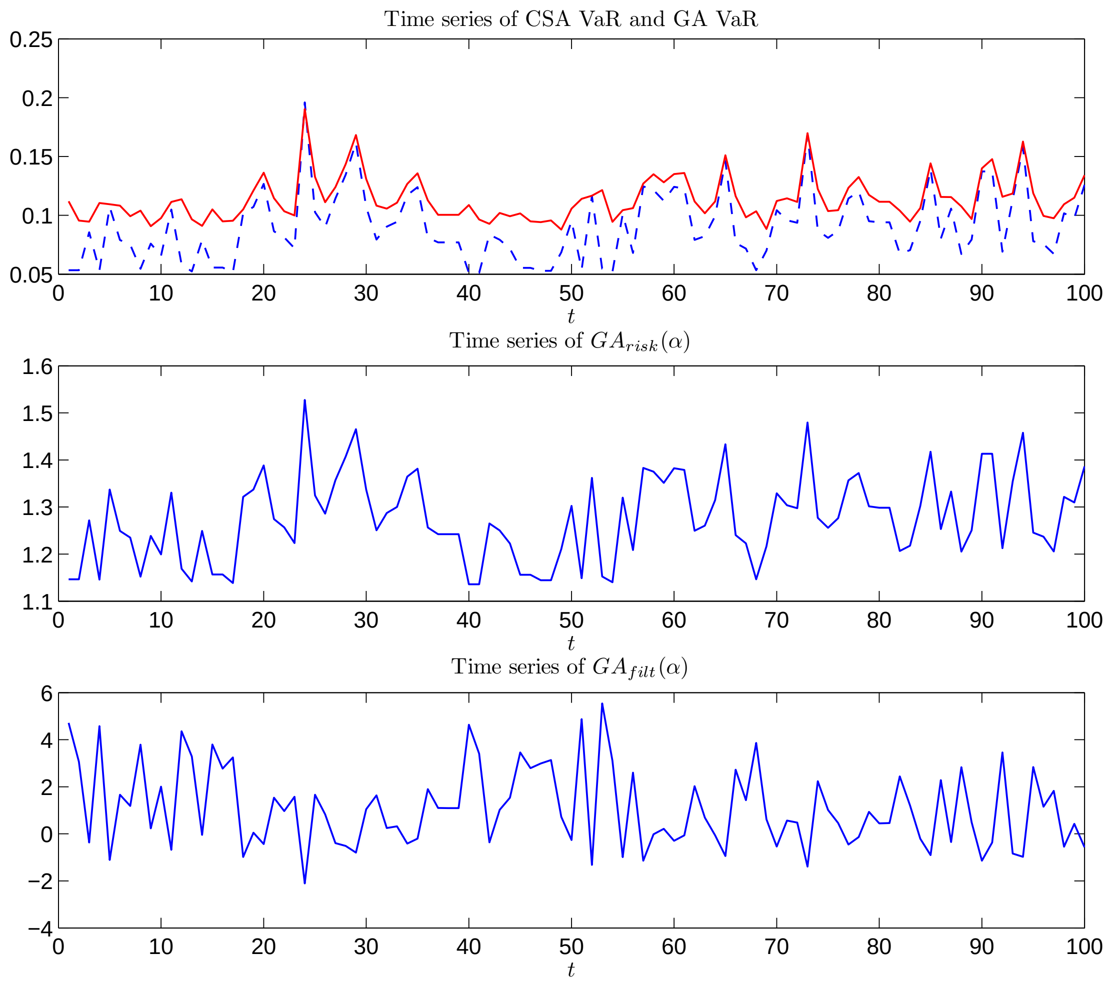
The left Panel displays the CSA VaR (dashed line), the GA VaR for $n = 100$ (solid line) and the GA VaR for $n = 1000$ (dotted line) as functions of the cross-sectional factor approximation $\hat{f}_{n,t}$. The middle and right Panels display the GA component for risk, and the GA component for filtering, respectively. The information set is such that $n_t/n = PD$ and $\hat{f}_{n,t-1} = \mu$. The confidence level is $1 - \alpha = 0.995$. The structural parameters are such that $ELGD = 0.45$, $PD = 5\%$, $\rho = 0.12$ and $\gamma = 0.5$. In particular, the unconditional factor mean is $\mu = 3.05$ (see Table 1).

Figure 6.9: Time series of simulated default frequencies, portfolio losses, systematic factors and cross-sectional approximations of the factor.



The upper and middle Panels display a simulated time series of default frequencies and percentage portfolio losses, respectively. The lower Panel displays the corresponding time series of factor values (circles) and cross-sectional factor approximations (squares). The portfolio size is $n = 100$. The structural parameters are such that $ELGD = 0.45$, $PD = 5\%$, $\rho = 0.12$ and $\gamma = 0.5$. In particular, the unconditional factor mean is $\mu = 3.05$ (see Table 1).

Figure 6.10: Time series of simulated CSA VaR, GA VaR, and GA risk and filtering components.



The upper Panel displays a simulated time series of CSA VaR (dashed line) and GA VaR (solid line) for portfolio size $n = 100$ and confidence level $1 - \alpha = 0.995$. The middle and lower Panels display the corresponding time series of GA risk and filtering components. The structural parameters are such that $ELGD = 0.45$, $PD = 5\%$, $\rho = 0.12$ and $\gamma = 0.5$ (see Table 1).

R E F E R E N C E S

Acerbi, C., and D., Tasche (2002): "On the Coherency of Expected Shortfall", Journal of Banking and Finance, 26, 1487-1503.

Bahadur, R. (1966): "A Note on Quantiles in Large Samples", Annals of Mathematical Statistics, 37, 577-580.

Basel Committee on Banking Supervision (1995): "An Internal Model Based Approach to Market Risk Capital Requirements", Bank of International Settlements, April, 1995.

Basel Committee on Banking Supervision (2001): "The New Basel Capital Accord", Consultative Document of the Bank for International Settlements, April 2001, Part 2: Pillar 1.

Basel Committee on Banking Supervision (2003): "The New Basel Capital Accord", Consultative Document of the Bank for International Settlements, April 2003, Part 3: The Second Pillar.

Gagliardini, P., and C., Gourieroux (2010): "Granularity Adjustment for Risk Measures: Systematic and Unsystematic Risks", CREST DP.

Gagliardini, P., Gourieroux, C., and A., Monfort (2010): "Microinformation, Nonlinear Filtering and Granularity", forthcoming in Journal of Financial Econometrics.

Geske, R. (1977): "The Valuation of Corporate Liabilities as Compound Options", Journal of Financial and Quantitative Analysis, 12, 541-552.

Gordy, M. (2003): "A Risk-Factor Model Foundation for Ratings Based Bank Capital Rules", Journal of Financial Intermediation, 12, 199-232.

Gordy, M. (2004): "Granularity Adjustment in Portfolio Credit Risk Measurement", in *Risk Measures for the 21st Century*, Giorgio Szego ed., Wiley.

Gouriéroux, C. (2010): "Risk Measures: Statistical Estimation", Encyclopedia of Quantitative Finance, R. Cont Ed., Wiley.

Gouriéroux, C., Laurent, J.-P., and O., Scaillet (2000): "Sensitivity Analysis of Value-at-Risk", Journal of Empirical Finance, 7, 225-245.

Koenker, R. (2005): *Quantile Regression*, Cambridge University Press.

Martin, R., and T., Wilde (2003): "Unsystematic Credit Risk", Risk Magazine, 15, 123-128.

Merton, R. (1974): "On the Pricing of Corporate Debt: The Risk Structure of Interest Rates", Journal of Finance, 29, 449-470.

Pykhtin, M., and A., Dev (2002): "Analytical Approach to Credit Risk Modelling", Risk, 15, 26-32.

Tasche, D. (2000): "Conditional Expectation as Quantile Derivative", DP Bundesbank.

Vasicek, O. (1991): "Limiting Loan Loss Probability Distribution", DP KMV Corporation.

Wang, S. (1996): "Premium Calculation by Transforming the Layer Premium Density", ASTIN Bulletin, 26, 71-92.

Wang, S. (2000): "A Class of Distortion Operators for Pricing Financial and Insurance Risks", Journal of Risk and Insurance, 67, 15-36.

Wilde, T. (2001): "Probing Granularity", Risk, 14, 103-106.

Chapter 7

A. Review on Econometrics

A.1 Simulation

A.2 Efficiency Bounds

A.3 Panel Models

A.4 Singular Value Decomposition and Principal Component Analysis

A.5 Linear Prediction and Kalman Filter

A.6 The Newton-Raphson Algorithm

A.1 Simulation

Simulations are artificial data randomly drawn by the econometrician. Simulation based approaches are used to compute numerically complicated integrals (Monte-Carlo integration) and in particular derivative prices, to derive the finite sample properties of an estimator (e.g. bootstrap), or even to define new estimation methods (simulation based method of moments and indirect inference).

A.1.1 The principle

All simulation techniques are based on the following lemma:

Lemma A.1: *Let X be a one-dimensional random variable with continuous distribution and a strictly increasing cumulative distribution function (cdf) F. Then the variable $U = F(X)$ follows a uniform distribution on* $[0, 1]$.

Proof: Indeed, we have:

$$\begin{aligned} P[U \leq u] &= P[F(X) \leq u] \\ &= P[X \leq F^{-1}(u)] \quad \text{(since a continuous increasing function is invertible)} \\ &= F[F^{-1}(u)] = u, \end{aligned}$$

which is the cdf of the uniform distribution on $[0, 1]$.

QED

The lemma above implies the following corollary:

Corollary A.2: *Let X be a continuous variable with increasing cdf and Φ the cdf of the standard normal distribution, then the variable $\varepsilon = \Phi^{-1}[F(X)]$ is standard normal.*

The previous results can be directly used for simulating an artificial independent sample from distribution F, by using a software to produce i.i.d standard normal observations (rndn software), or i.i.d uniform observations (rndu software). The approach is for instance the following:

i) Draw at random S artificial data $\varepsilon_1, \dots, \varepsilon_S$ from the standard normal by the software rndn.

ii) Then compute the simulated values $X_1, \dots, X_S$ by $X_s = F^{-1}[\Phi(\varepsilon_s)]$.

This approach is easily extended to multivariate random variables. As an illustration, let us consider a bivariate vector (X, Y) with known distribution. This distribution is characterized by the marginal distribution of X with cdf $F_X(x)$ and the conditional distribution of Y given $X = x$, with conditional cdf $F_{Y|X}(y|x)$.

The simulation approach is the following:

i) Draw at random two independent samples of size S from the standard normal by software rndn. These samples are ε_s, η_s, for $s = 1, \dots, S$.

ii) The simulated values $X_1, \dots, X_S$ are computed as:

$$X_s = F_X^{-1}[\Phi(\varepsilon_s)], \quad s = 1, \dots, S.$$

iii) Then the simulated values $Y_1, \dots, Y_S$ are deduced by:

$$Y_s = F_{Y|X}^{-1}[\Phi(\eta_s)|X_s], \quad s = 1, \dots, S.$$

where $F_{Y|X}^{-1}(.|x)$ is the inverse of the conditional cdf with respect to argument y.

The simulation scheme is completely fixed by the analyst, who has to choose the form of the distribution, but also the number of replications.

A.1.2 Monte-Carlo integration

Integrals, or equivalently expectations, can be computed by simulation. Let us consider an expectation:

$$E_0(Y) = \int y p_0(y) dy, \tag{A.1.1}$$

where p_0 is a known probability density function. Then, we can draw S independent observations $Y_1, \ldots, Y_S$ from distribution p_0. By the Law of Large Numbers, the sample mean of these simulated values $\bar{Y}_S = \frac{1}{S}\sum_{s=1}^{S} Y_s$ is a consistent approximation of the true unknown expectation as $S \to \infty$, that is, $\bar{Y}_S \simeq E_0(Y)$ for large S.

This approach can be extended to improve the accuracy of the approximation. Let us introduce another given p.d.f. $q_0(y)$, called the **importance function**. Then, we have:

$$E_0(Y) = \int y \frac{p_0(y)}{q_0(y)} q_0(y) dy. \tag{A.1.2}$$

An approximation of the expectation can be derived as follows :

i) Draw at random S observations $Y_1^*, \ldots, Y_S^*$ from the distribution q_0.

ii) Then approximate the expectation by :

$$\frac{1}{S}\sum_{s=1}^{S} [Y_s^* p_0(Y_s^*)/q_0(Y_s^*)].$$

This approximation is very accurate when $q_0(y)$ is almost proportional to $y p_0(y)$.

A.1.3 Bootstrap

We can now explain how to derive the properties of an estimator for a large, but finite number of observations. As an illustration, let us consider a parametric dynamic model:

$$y_t = a(y_{t-1}, \varepsilon_t, \theta), \quad t = 1, \ldots, T,$$

where ε_t are iid standard Gaussian variables and a is a known function. Let us denote $\hat{\theta}_T = \hat{\theta}(y_1, \ldots, y_T)$ a consistent estimator of parameter θ. Estimator $\hat{\theta}_T$ is a good approximation of θ, which can be used to simulate several artificial paths for Y. More precisely, we can:

i) Draw a sequence of size T from the standard normal distribution. This sequence is denoted by $\varepsilon_1^s, \ldots, \varepsilon_T^s$.

ii) Deduce the simulated path by recursion:

$$y_t^s = a(y_{t-1}^s, \varepsilon_t^s, \hat{\theta}_T), \quad t = 1, \ldots, T \text{ (with } y_0^s = y_0).$$

iii) Compute the simulated estimate from this path, as :

$$\hat{\theta}_T^s = \hat{\theta}(y_1^s, \ldots, y_T^s).$$

iv) Replicate the approach for $s = 1, \ldots, S$.

v) Then, for large S, the sample distribution of $(\hat{\theta}_T^1, \ldots, \hat{\theta}_T^S)$ is a good approximation of the unknown distribution of the estimator.

Further reading

Caflish, R. (1998): "Monte-Carlo and Quasi Monte-Carlo Methods", Acta Numerica, Vol 7, Cambridge University Press, p1-49.

Efron, B., and R., Tibshirani (1994): "An Introduction to the Bootstrap", Chapman & Hall.

A.2 Efficiency Bounds

The accuracy of an estimator $\hat{\theta}$ of a (multidimensional) parameter θ depends on its bias, that is, the difference between the expectation of $\hat{\theta}$ and the true parameter value, and on its variance-covariance matrix.

For a consistent estimator, the bias is asymptotically equal to zero and its accuracy is entirely captured by its asymptotic variance-covariance matrix. It is often possible to find a lower bound for the asymptotic variance-covariance matrix of the consistent estimators of θ. This bound, when it exists, is called an (asymptotic) **efficiency bound**.

Then, a consistent estimator, whose variance-covariance matrix coincides asymptotically with the efficiency bound, is preferable to any other consistent estimator. It is called an (asymptotically) **efficient estimator**.

A.2.1 Parametric model parametrized by θ

Let us first consider a parametric model with likelihood function $l_n(y;\theta)$, where y denotes the vector of observations and n their number. An efficiency bound is given by:

$$B(\theta) = [I(\theta)]^{-1}, \tag{A.2.1}$$

where the information matrix $I(\theta)$ can be approximated by:

$$I(\theta) \simeq E_\theta \left[-\frac{\partial^2 \log l_n(y;\theta)}{\partial\theta\partial\theta'} \right], \tag{A.2.2}$$

and E_θ denotes the expectation computed with the value θ of the parameter. In this framework $B(\theta)$ is called the **parametric efficiency bound**.

Under standard regularity conditions, the maximum likelihood estimator:

$$\hat{\theta} = \arg\max_\theta \log l_n(y;\theta)$$

has an (asymptotic) variance-covariance matrix equal to the (asymptotic) efficiency bound. Thus, the maximum likelihood estimator is (asymptotically) parametrically efficient.

A.2.2 Parametric model partly parametrized by θ

Let us now consider a parametric model including also nuisance parameters β, say. The likelihood function is $l_n(y;\theta,\beta)$, and the parametric efficiency bound for the whole parameter $(\theta',\beta')'$ is:

$$B(\theta,\beta) = [I(\theta,\beta)]^{-1}, \tag{A.2.3}$$

where

$$I(\theta,\beta) \simeq E_{\theta,\beta}\left[-\frac{\partial^2 \log l_n(y;\theta,\beta)}{\partial\begin{pmatrix}\theta\\ \beta\end{pmatrix}\partial\begin{pmatrix}\theta\\ \beta\end{pmatrix}'}\right]. \tag{A.2.4}$$

This information matrix can be decomposed into blocks as:

$$I(\theta,\beta) = \begin{pmatrix} I_{\theta\theta} & I_{\theta\beta} \\ & \\ I_{\beta\theta} & I_{\beta\beta}\end{pmatrix}, \quad \text{say,}$$

where:

$$\begin{aligned}
I_{\theta\theta} &= E_{\theta,\beta}\left[-\frac{\partial^2 \log l_n(Y;\theta,\beta)}{\partial\theta\partial\theta'}\right], \\
I_{\theta\beta} &= E_{\theta,\beta}\left[-\frac{\partial^2 \log l_n(Y;\theta,\beta)}{\partial\theta\partial\beta'}\right] = I'_{\beta,\theta}, \\
I_{\beta\beta} &= E_{\theta,\beta}\left[-\frac{\partial^2 \log l_n(Y;\theta,\beta)}{\partial\beta\partial\beta'}\right].
\end{aligned}$$

By block inversion, we deduce the North-West block of the efficiency bound, which provides the parametric efficiency bound for θ in presence of a nuisance parameter β. It is given by:

$$B_{\theta\theta}(\theta,\beta) = [I_{\theta\theta} - I_{\theta\beta}(I_{\beta\beta})^{-1}I_{\beta\theta}]^{-1}. \tag{A.2.5}$$

A.2.3 Semi-parametric efficiency bound

Let us now consider a semi-parametric model, in which the likelihood function is parametrized by a parameter of interest θ and by a functional nuisance parameter $g \in G$. While the parameter of interest is a standard (finite-dimensional) vector, the nuisance parameter can be an infinite-dimensional object, e.g. the unknown density function of the error in a regression model. This semi-parametric model nests all parametric models in which function g has been parametrized $g = g_\beta$, say. The parametric efficiency bound for such a nested parametric model $B_{\theta\theta}(g_\beta)$ can be computed by equation (A.2.5), and depends on the chosen parametrization of function g. The **semi-parametric efficiency bound** is defined as the maximal (i.e. the least favorable) bound corresponding to all admissible nested parametric models:

$$B_{\theta\theta}(\theta, g) = \max_{g_\beta} B_{\theta\theta}(g_\beta). \tag{A.2.6}$$

A consistent estimator is **semi-parametrically efficient**, if its variance-covariance matrix is asymptotically equal to the semi-parametric efficiency bound. An example of semi-parametrically efficient estimator is the Ordinary Least Squares (OLS) estimator in a linear regression model under the standard regularity assumptions on the errors.

Further reading

Chamberlain, G. (1992): "Efficiency Bounds for Semiparametric Regressions", Econometrica, 60, 567-596.

Cramer, H. (1946): "Mathematical Methods of Statistics", Princeton Univ. Press.

Rao, R. (1945): "Information and the Accuracy Attainable in the Estimation of Statistical Parameters", Bulletin of the Calcutta Mathematical Society, 37, 81-89.

A.3 Panel Models

Panel models are explanatory models for panel data, that are observations $y_{i,t}$, $i = 1, \ldots, n$, $t = 1, \ldots, T$ doubly indexed by individual and time. The basic Gaussian linear model for panel data is:

$$y_{i,t} = \alpha + x'_{i,t} b + w_{i,t}, \quad i = 1, \ldots, n, \quad t = 1, \ldots, T, \tag{A.3.1}$$

where $x_{i,t}$ are the observations of the explanatory variables and $\omega_{i,t}$ are independent, Gaussian error terms with common distribution $N(0, \sigma_\omega^2)$.

The basic model (A.3.1) is usually extended to highlight possible individual, or time effects. These effects can be assumed either fixed, or random.

A.3.1 Panel model with fixed effects

The introduction of fixed effects leads to the model:

$$y_{i,t} = \alpha + \beta_i + \gamma_t + x'_{it} b + \omega_{it}, i = 1, \ldots, n, t = 1, \ldots, T, \tag{A.3.2}$$

where β_i andγ_t are additional parameters satisfying the constraints:

$$\beta. = \sum_{i=1}^{n} \beta_i = 0, \quad \gamma. = \sum_{t=1}^{T} \gamma_t = 0,$$

to avoid collinearity problems. Parameters β_i (resp. γ_t) are the fixed individual effects (resp. time effects). Model (A.3.2) is a special case of linear model and the parameters α, β_i, for $i = 1, \ldots, n$, and γ_t, for $t = 1, \ldots, T$, can be estimated by Ordinary Least Squares (OLS). However, the OLS estimators do not feature standard asymptotic properties. The reason is that the total number of parameters equal to $n + T - 1$ is not fixed, but increases with the number of observations, that is, with n and T. This is the so-called **incidental parameter problem**.

The OLS estimators have closed form expressions, which are easily interpreted for a fixed effect model without explanatory variable x, that is,

$$y_{i,t} = \alpha + \beta_i + \gamma_t + \omega_{i,t}, \tag{A.3.3}$$

with $\omega_{i,t} \sim IIN(0, \sigma_\omega^2)$. Let us define the following sample means:

$$\bar{\bar{y}} = \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} y_{i,t}, \quad \bar{y}_{i\cdot} = \frac{1}{T} \sum_{t=1}^{T} y_{i,t}, \quad \bar{y}_{\cdot t} = \frac{1}{n} \sum_{i=1}^{n} y_{i,t}. \tag{A.3.4}$$

The OLS estimators of the parameters are:

$$\hat{\alpha} = \bar{\bar{y}}, \quad \hat{\beta}_i = \bar{y}_{i\cdot} - \bar{\bar{y}}, \quad \hat{\gamma}_t = \bar{y}_{\cdot t} - \bar{\bar{y}}, \tag{A.3.5}$$

whereas the residuals are given by:

$$\hat{\omega}_{i,t} = y_{i,t} - \bar{y}_{i\cdot} - \bar{y}_{\cdot t} + \bar{\bar{y}}. \tag{A.3.6}$$

Hence, the estimate of the constant is the full sample average of the observations across individual and time, while the estimates of the individual (resp. time) effects are the differences between the time (resp. individual) averages and the full sample average.

A.3.2 Panel model with random effects

In this extension, the individual and time effects are assumed stochastic. The model becomes:

$$y_{i,t} = \alpha + x_{i,t}' b + u_i + v_t + \omega_{i,t}, \quad i = 1, \ldots, n, \quad t = 1, \ldots, T, \tag{A.3.7}$$

where $u_i, v_t, \omega_{i,t}$ are independent Gaussian variables, independent of the explanatory variables, with distributions:

$$u_i \sim N(0, \sigma_u^2), \quad v_t \sim N(0, \sigma_v^2), \quad \omega_{i,t} \sim N(0, \sigma_\omega^2),$$

respectively.

In this extension, the number of parameters is fixed, which solves the incidental parameter problem. However, this linear model has a non scalar variance-covariance matrix function of the three parameters $\sigma_u^2, \sigma_v^2, \sigma_\omega^2$. Except in very special cases (see Chapter 2), the maximum likelihood estimators of the parameter do not admit closed form expressions. Moreover, they have nonstandard asymptotic properties, which depend on the assumed asymptotics, either $n \to \infty$ and T fixed, or n fixed and $T \to \infty$, or $n \to \infty$ and $T \to \infty$.

A.3.3 Panel model with both fixed and random effects

It is not possible to introduce in a panel model a fixed and a random effect of a same type, since the fixed effect will systematically capture the random effect. Thus, there exist only four possibilities for a panel model with both individual and time effects as seen in the table below.

Table A.1: Panel models with both individual and time effects

	time effect	
individual effect	fixed	random
fixed	X	X
random	X	X

A.3.4 Fixed or random effects

There exist testing procedures for choosing between fixed and random effects in panel models. However, it is often preferable to base this choice according to the problem of interest. This choice is well illustrated by applications to credit. Let us assume that y_{it} is a quantitative measure of individual risk. The models with fixed individual effect,

$$y_{it} = \alpha + \beta_i + x_{it}b + w_{it},$$

are used in a first step to make a segmentation of the set of contracts into rather homogenous segments. This segmentation is done as follows: first, estimate the individual fixed effects $\hat{\beta}_i$, $i = 1, \dots, n$. Second, define the segments from these values $\hat{\beta}_i$ by an appropriate discretization. Segment k, with $k = 1, \dots, K$, includes the individuals i such that $a_{k-1} < \hat{\beta}_i \leq a_k$, where a_k, $k = 0, \dots, K$, is a given set of thresholds.

In the current Basel regulation, such a segmentation has to be defined before analyzing more precisely the risks within and between segments. This second step of the risk analysis has to account for the possible dependencies between two individuals risks at a same time, or the successive risks of a same individual at two different times. This is done by introducing in each segment a random time effect (resp. a random individual effect), since by definition fixed effects are deterministic and thus non risky.

Further reading

Arellano, M., and B., Honore (2001): "Panel Data Models : Some Recent Developments", Handbook of Econometrics, ed. J., Heckman and E. Leamer, Chap 53, p 3231-3296.

Baltagi, B. (1995): "Econometric Analysis of Panel Data", Wiley.

A.4 Singular Value Decomposition and Principal Component Analysis

Principal Component Analysis (PCA) is based on the analysis of eigenvalues and eigenvectors of well-chosen symmetric matrices. We first recall basic decompositions of matrices in linear algebra.

A.4.1 Singular Value Decomposition (SVD)

i) Spectral decomposition of a symmetric matrix

Any symmetric matrix Ω of dimension (n, n) can be diagonalized. This matrix admits real eigenvalues λ_i, $i = 1, \ldots, n$, and real eigenvectors u_i, $i = 1, \ldots, n$. These eigenvectors form an orthonormal basis. They can always be chosen such that $\langle u_i, u_j \rangle = u_i' u_j = 0$, if $i \neq j$, and $\|u_i\|^2 = u_i' u_i = 1$, $\forall i$.

Let us denote by Q the matrix whose columns are these eigenvectors. The orthonormality restrictions imply $Q^{-1} = Q'$ and the matrix Ω can be written as:

$$\Omega = Q\Lambda Q', \tag{A.4.1}$$

where Λ is the diagonal matrix with the eigenvalues of Ω as diagonal elements. Equivalently, the equality (A.4.1) can be written as:

$$\Omega = \sum_{i=1}^{n} \lambda_i u_i u_i', \tag{A.4.2}$$

which gives the **spectral decomposition** of matrix Ω.

ii) SVD of a rectangular matrix

Let us now consider a rectangular matrix X with dimension (n, T). Typically, X can be a matrix of observations doubly indexed by individual and time. This matrix can be used to construct two symmetric matrices by considering the squared matrices XX' and $X'X$. These matrices are symmetric positive semi-definite with respective sizes (n, n) and (T, T).

Then, we can consider their spectral decompositions:

$$XX' = \sum_{i=1}^{n} \lambda_i u_i u_i', \quad \lambda_i \geq 0, \tag{A.4.3}$$

$$X'X = \sum_{t=1}^{T} \mu_t v_t v_t', \quad \mu_t \geq 0. \tag{A.4.4}$$

The following lemma explains that these spectral decompositions can be chosen strongly linked.

Lemma A.3: *Let us denote* $K = \min(n, T)$

i) *We can order the eigenvalues such that* $\lambda_k = \mu_k$, $k = 1, \ldots, K$, *the remaining eigenvalues being equal to zero.*

ii) *The two orthonormal basis can be chosen such that:*

$$\langle v_k, u_k \rangle = 1, \quad k = 1, \ldots, K, \quad \langle v_k, u_j \rangle = 0, \quad \forall k \neq j = 1, \ldots K.$$

iii) *Matrix X can be decomposed as:*

$$X = \sum_{k=1}^{K} \sqrt{\lambda_k} u_k v_k'.$$

Thus, we have the following decompositions:

$$\begin{cases} X = U\Lambda^{1/2}V', \\ \\ XX' = U\Lambda U', \\ \\ X'X = V\Lambda V', \end{cases} \tag{A.4.5}$$

where Λ is the diagonal matrix with elements λ_k, $k = 1, \ldots, K$, U (resp. V) is the matrix with columns u_k, $k = 1, \ldots, K$ (resp. v_k, $k = 1, \ldots, K$) and $\Lambda^{1/2}V = X'U$.

The above decomposition of matrix X is the **singular value decomposition** of X; the vectors u_k (resp. v_k) are its **left singular vectors** (resp. **right singular vectors**), and $\lambda_k^{1/2}$ **the singular values**.

A.4.2 Principal Component Analysis

When the eigenvalues $\lambda_k = \mu_k$, $k = 1, \ldots, K$, are different, these eigenvalues can be ranked in decreasing order: $\lambda_1 > \lambda_2 \ldots > \lambda_K$. The Principal Component Analysis (PCA) proposes interpretations of these eigenvalues and of the associated eigenvectors.

To understand the PCA interpretation, let us consider the constrained optimisation problem:

$$\begin{cases} \max_{a \in \mathbb{R}^n} a'XX'a \\ \\ s.t. \ a'a = 1, \end{cases} \tag{A.4.6}$$

and introduce a Lagrange multiplier ν. The corresponding Lagrangean function $a'XX'a - 2\nu(a'a-1)$ can be optimized with respect to vector a. The first-order condition is:

$$XX'a^* - \nu a^* = 0 \Leftrightarrow XX'a^* = \nu a^*,$$

and the optimal value of the objective function in (A.4.6) is:

$$a^{*'}XX'a^* = a^{*'}(\nu a^*) = \nu. \tag{A.4.7}$$

Equation (A.4.7) means that the solution a^* is an eigenvector of matrix XX', whereas the associated eigenvalue ν, equal to the value of the objective function, has to be maximized. Thus, the Lagrange multiplier is equal to the largest eigenvalue $\nu^* = \lambda_1 = \mu_1$, and the solution of the optimization problem (A.4.6) is the normalized eigenvector u_1. By using the orthogonality between the eigenvectors u_1, u_2, ..., u_K, such optimization can be performed in a recursive way as described in the following property:

Property A.4:
i) *Eigenvector u_1 is solution of the optimization problem:*

$$\max_a a'XX'a,\ s.t.\ a'a = 1,$$

whereas λ_1 is the associated value of the objective function.

ii) *Eigenvector u_2 is solution of the optimization problem:*

$$\max_a a'XX'a,\ s.t.\ a'a = 1, \textit{and } a'u_1 = 0,$$

whereas λ_2 is the associated value of the objective functions, and so on.

The property above is usually applied to a square matrix XX' interpretable as a variance-covariance matrix. Let us consider panel data y_{it}, $i = 1, \dots, n$, $t = 1, \dots, T$, and let X be the (n, T) matrix with elements $x_{i,t} = y_{it} - \bar{y}_{i\cdot}$. Then, the matrix $\frac{1}{T}XX'$ is simply the sample variance-covariance matrix of variables y_i with observations $y_{i1}, \ldots, y_{iT}$. The associated eigenvectors $u_1, u_2 \ldots$, called **principal components**, provide the directions, which are the most variable, the second most variable... For instance, if $y_{it} = r_{it}$ has the interpretation of an asset return, $u_1 = (u_{11}, \ldots, u_{1n})'$ can be interpreted as a portfolio allocation corresponding to the most risky portfolio allocations (under the constraint $a'a = 1$). The demeaned values of the return of this portfolio are equal to $\sum_{i=1}^{n} u_{1i}(y_{it} - \bar{y}_{i\cdot})$, $t = 1, \ldots, T$. Since $\sqrt{\lambda_1} v_1 = X' u_1$, these portfolio returns are equal to the components of the first eigenvector of $X'X$ scaled by $\sqrt{\lambda_1}$.

Further reading

Gantmacher, F. (1959): "Theory of Matrices", Vol 1 and 2, American Mathematical Society Chelsea Publishing, Providence.

Jolliffe, I. (2002): "Principal Component Analysis", 2nd ed, Springer, New-York.

Pearson, K. (1901): "On Lines and Planes of Closest Fit to Systems of Points in Space", Philosophical Magazine, 2, 559-572.

A.5 Prediction and Kalman Filter

A.5.1 Linear Prediction

Let us consider two random vectors X and Y with dimensions K and n, respectively, with means $E(X) = m_X$, $E(Y) = m_Y$, variances $V(X) = \Sigma_{XX}$, $V(Y) = \Sigma_{YY}$, and cross-covariances $Cov(X, Y) = \Sigma_{XY}$, $Cov(Y, X) = \Sigma_{YX}$. The mean vectors m_X, m_Y have dimensions $(K, 1)$ and $(n, 1)$, respectively.

Matrix Σ_{XX} (resp. Σ_{YY}) is the variance-covariance matrix of X with dimension (K, K) [resp. of Y with dimension (n, n)]. The cross-covariances are such that $\Sigma_{XY} = \Sigma'_{YX}$ has dimension (K, n).

The linear prediction of vector Y based on X is a vector $\hat{Y} = \hat{A}X + \hat{b}$ such that:

$$(\hat{A}, \hat{b}) = \arg\min_{A,b} E\left[\|Y - AX - b\|^2\right].$$

Thus, each component $\hat{Y}_i$ provides the best linear approximation of Y_i based on $X_1, \ldots, X_K$ with possibly an intercept. The expression of the linear prediction and of the prediction error are given below.

Property A.5: i) *Let us assume Σ_{XX} invertible. The best linear prediction of Y based on X is:*

$$\hat{Y} = m_Y + \Sigma_{YX}(\Sigma_{XX})^{-1}(X - m_X).$$

ii) *The prediction error $\hat{u} = Y - \hat{Y}$ is zero-mean: $E(\hat{u}) = 0$, with a variance-covariance matrix:*

$$V(\hat{u}) = \Sigma_{YY} - \Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}.$$

When X, Y are jointly Gaussian, the linear prediction $\hat{Y}$ coincides with the conditional expectation of Y given X, denoted $E(Y|X)$, and the residual variance $V(\hat{u})$ with the conditional variance-covariance matrix $V(Y|X)$.

A.5.2 Kalman Filter

The standard Gaussian linear state space model assumes:

State equation: $F_t = \Phi F_{t-1} + \eta_t$

Measurement equation: $y_t = BF_t + \varepsilon_t$, where y_t (resp. F_t) has dimension n (resp. K), the errors η_t, ε_t are independent Gaussian white noises $\eta_t \sim N(0, \Omega_\eta)$, $\varepsilon_t \sim N(0, \Omega_\varepsilon)$. The matrices Φ, B, Ω_ε, Ω_η are assumed given.

The Kalman filter is a set of algorithms to compute recursively (linear) predictions of F_t and y_t, and their accuracy. These linear predictions can be either of the type $E(F_t|y_t, \ldots, y_0)$, $E(y_t|y_{t-1}, \ldots, y_0)$, or of the type $E(F_t|y_T, \ldots, y_0)$. When factor F_t is approximated by current and lagged observed values, the algorithm is called a **filter**. When the information includes also future values, it is called a **smoother**.

The filter and smoother algorithms have been derived by Kalman, using previous results by Thiele and Swerling. They are based on a recursive use of the linear prediction formula in Property A.5. Let us denote:

$$
\begin{aligned}
\hat{F}_{t|t} &= E(F_t|\underline{y_t}), \text{ where } \underline{y_t} = (y_t, y_{t-1}, \ldots), \\
\hat{F}_{t|t-1} &= E(F_t|\underline{y_{t-1}}), \\
\Sigma_{t|t} &= V(F_t|\underline{y_t}), \quad \Sigma_{t|t-1} = V(F_t|\underline{y_{t-1}}).
\end{aligned}
$$

The filter involves the following recursions:

Prediction:

Predicted factor: $\hat{F}_{t|t-1} = \Phi \hat{F}_{t-1|t-1}$,

Accuracy of the predicted factor: $\Sigma_{t|t-1} = \Phi \Sigma_{t-1|t-1} \Phi' + \Omega_\eta$.

Updating:

Measurement residual: $\hat{u}_{t|t} = y_t - B\hat{F}_{t|t-1}$,

Residual variance: $H_{t|t} = B\Sigma_{t|t-1}B' + \Omega_\varepsilon$,

Kalman gain: $K_{t|t} = \Sigma_{t|t-1}B'(H_{t|t})^{-1}$, Updated predicted factor: $\hat{F}_{t|t} = \hat{F}_{t|t-1} + K_{t|t}\hat{u}_{t,t}$,

Updated accuracy of the predicted factor: $\Sigma_{t|t} = (Id - K_{t|t}B)\Sigma_{t|t-1}$.

Further reading

Harvey, A. (1989): "Forecasting, Structural Time Series Models and the Kalman Filter", Cambridge University Press.

Kalman, R. (1960): "A New Approach to Linear Filtering and Prediction, Problems", Journal of Basic Engineering, 82, 35-45.

Lauritzen, S. (2002): "Thiele: Pioneer in Statistics", Oxford University Press.

A.6 The Newton-Raphson Algorithm

i) The basic algorithm

This is the best known method to find numerically the solutions of a nonlinear system of equations. The modern presentation of the algorithm is due to T. Simpson (1740), based on earlier works by the Persian mathematician Sharaf al-Din al-Tusi (1135-1213), I. Newton (1669) and J. Raphson (1690).

Let us consider a differentiable function g from $I\!R^P$ to $I\!R^P : \theta \to g(\theta)$, say, and denote $\frac{\partial g}{\partial \theta'}(\theta)$ its gradient, with elements the different partial derivatives. The idea of the algorithm is to replace the initial nonlinear system:

$$g(\theta) = 0, \tag{A.6.1}$$

by its linear expansion (first-order Taylor approximation) around some value θ_0:

$$g(\theta_0) + \frac{\partial g}{\partial \theta'}(\theta_0)(\theta - \theta_0) = 0, \tag{A.6.2}$$

whose solution has an explicit form:

$$\theta = \theta_0 - \left[\frac{\partial g}{\partial \theta'}(\theta_0)\right]^{-1} g(\theta_0). \tag{A.6.3}$$

The solution in (A.6.3) is well-defined if matrix $\frac{\partial g}{\partial \theta'}(\theta_0)$ is non-singular, that is, function g is one-to-one locally around θ_0.

The Newton-Raphson algorithm applies this approach iteratively, along the following steps:

Step 1: Choose a starting value $\theta^{(0)}$.

Step 2: Then apply recursively formula (A.6.3):

$$\theta^{(p+1)} = \theta^{(p)} - \left[\frac{\partial g}{\partial \theta'}(\theta^{(p)})\right] g(\theta^{(p)}).$$

Step 3: Stop when numerical convergence is reached.

ii) Application to maximum likelihood

In the standard cases, the maximum likelihood estimate is solution of the first-order conditions:

$$\frac{\partial \log l(y;\hat{\theta})}{\partial \theta} = 0, \tag{A.6.4}$$

where l is the joint likelihood function of observations y, vector θ is the parameter, and $\hat{\theta}$ its maximum likelihood estimate.

Then, the recursive equation of the Newton-Raphson algorithm becomes:

$$\theta^{(p+1)} = \theta^{(p)} + \left(-\frac{\partial^2 \log l(y;\theta^{(p)})}{\partial \theta \partial \theta'}\right)^{-1} \frac{\partial \log l(y;\theta^{(p)})}{\partial \theta}. \tag{A.6.5}$$

When the numerical convergence is reached, we get a solution $\hat{\theta}$ of the likelihood equations, which may be the ML estimate. If it is the case, the quantity:

$$\left[-\frac{\partial^2 \log l(y;\hat{\theta})}{\partial \theta \partial \theta'}\right]^{-1}$$

involved in the recursive equation provides the estimated variance-covariance matrix of the maximum likelihood estimator.

The choice of the starting value can accelerate significantly the algorithm. In particular, we have following the result:

Property A.6: *Let us consider a consistent estimator $\tilde{\theta}$ of parameter θ. Then*

$$\tilde{\theta}^{(1)} = \tilde{\theta} + \left[-\frac{\partial^2 \log l(y;\tilde{\theta})}{\partial\theta\partial\theta'} \right]^{-1} \frac{\partial \log l(y,\tilde{\theta})}{\partial\theta}$$

is convergent and asymptotically efficient.

Therefore, with this choice, a single iteration is enough.

Further reading

Ypma, T. (1995): "Historical Development of the Newton-Raphson Method", SIAM Review, 37, 531-551.

Chapter 8

B. Review on Financial Theory

B.1. Portfolio Management

B.2. Arbitrage

B.3 Risk Measures

B.1 Portfolio Management

B.1.1 Portfolio characteristics

We consider n risky assets and one riskfree asset. Their unitary prices at date t are $p_{i,t}, i = 1, \ldots, n$ and 1, respectively, and their values at $t+1$ are $p_{i,t}(1 + r_{i,t+1}), i = 1, \ldots, n$, and $1 + r_{f,t}$, respectively, where $r_{i,t+1}$ is the return on asset i, and $r_{f,t}$ the riskfree return.

A **portfolio allocation** defines the quantity of each asset included in the portfolio. These quantities are denoted $a_{i,t}, i = 1, \ldots, n$, and $a_{0,t}$ at date t, where $a_{0,t}$ is the quantity in riskfree asset. The portfolio value at date t is :

$$W_t(\tilde{a}_t) = \sum_{i=1}^{n} a_{i,t} p_{i,t} + a_{0,t} = a_t' p_t + a_{0,t}, \tag{B.1.1}$$

with $a_t = (a_{1,t}, \ldots, a_{n,t})', \tilde{a}_t = (a_t', a_{0,t})', p_t = (p_{1,t}, \ldots, p_{n,t})'$.

Its value at date $t+1$ is :

$$\begin{aligned} W_{t+1}(\tilde{a}_t) &= \sum_{i=1}^{n} a_{i,t} p_{i,t}(1 + r_{i,t+1}) + a_{0,t}(1 + r_{f_t}) \\ &= W_t(\tilde{a}_t)(1 + r_{f,t}) + \sum_{i=1}^{n} a_{i,t} p_{i,t} r^*_{i,t+1}, \end{aligned} \tag{B.1.2}$$

where $r^*_{i,t+1} = r_{i,t+1} - r_{f,t}$ denotes the **excess return**. At date t, the allocation, the current prices and the riskfree rate are known, but the excess returns are unknown. Let us denote $y_{i,t+1} = p_{i,t} r^*_{i,t+1}, i = 1, \ldots, n$ the **excess gains** in the different risky assets. These excess gains are random at date t, with mean and variance given by:

$$\mu_t = E_t(y_{t+1}), \Sigma_t = V_t(y_{t+1}), \tag{B.1.3}$$

where E_t, V_t denote the expectation and variance-covariance matrix given the information available at date t.

The first and second-order conditional moments of the future portfolio value are:

$$
\begin{aligned}
E_t[W_{t+1}(\tilde{a}_t)] &= W_t(\tilde{a}_t)(1 + r_{f,t}) + a_t'\mu_t, &\text{(B.1.4)}\\
V_t[W_{t+1}(\tilde{a}_t)] &= a_t'\Sigma_t a_t, &\text{(B.1.5)}
\end{aligned}
$$

by using equation (B.1.2).

B.1.2 Mean-Variance portfolio management

In the mean-variance approach, the allocation is chosen to maximise a criterion taking into account the expected gain and the risk, under a budget constraint. More precisely, the optimization problem is:

$$
\begin{cases}
\max_{\tilde{a}} E_t[W_{t+1}(\tilde{a}_t)] - \dfrac{A}{2} V_t[W_{t+1}(\tilde{a}_t)] \\
\\
\text{s.t. } W_t(\tilde{a}_t) = W_{0,t},
\end{cases}
\qquad \text{(B.1.6)}
$$

where $A > 0$ is a measure of absolute risk aversion. The criterion is increasing in the expected portfolio value, decreasing in its variance, which creates a trade off between expected gain and risk.

The budget constraint and equations (B.1.4)-(B.1.5) can be used to deduce an unconstrained optimization problem in the allocation a_t in risky assets:

$$\max_{a_t} a_t'\mu_t - \frac{A}{2} a_t'\Sigma_t a_t. \qquad \text{(B.1.7)}$$

The first-order condition of problem (B.1.7) is:

$$\mu_t - A\Sigma_t a_t = 0. \qquad \text{(B.1.8)}$$

This provides the optimal allocation:

$$a_t^* = \frac{1}{A}\Sigma_t^{-1}\mu_t, \qquad \text{(B.1.9)}$$

called the **mean-variance efficient allocation**. The quantity invested in riskfree asset is then deduced from the budget constraint. We have:

$$a_{0,t}^* = W_{0,t} - a_t^{*'} p_t = W_{0,t} - \frac{1}{A}\mu_t' \Sigma_t^{-1} p_t. \tag{B.1.10}$$

B.1.3 The Sharpe performance

At the optimum, the criterion becomes :

$$W_{0,t}(1+r_{f,t}) + a_t^{*'}\mu_t - \frac{A}{2} a_t^{*'}\Sigma_t a_t^* = W_{0,t}(1+r_{f,t}) + \frac{1}{2A}\mu_t'\Sigma_t^{-1}\mu_t.$$

It depends on the stochastic properties of risky excess returns by means of the quantity :

$$S_t = \mu_t'\Sigma_t^{-1}\mu_t, \tag{B.1.11}$$

called **Sharpe performance** of the set of risky assets. This quantity is equal to :

$$S_t = \frac{[E_t(W_{t+1}^*) - W_{0,t}(1+r_{f,t})]^2}{V_t(W_{t+1}^*)}, \tag{B.1.12}$$

where W_{t+1}^* is the future value of the efficient portfolio. Thus, $S_t^{1/2}$ provides a measure of the maximal risk adjusted expected gain for a portfolio based on these n risky assets and riskfree asset.

Further reading

Lintner, J. (1965): "The Valuation of Risky Assets and the Selection of Risky Investments in Stock Portfolio and Capital Budgets", Review of Economic and Statistics, 47, 13-37.

Markowitz, H. (1952): "Portfolio Selection", The Journal of Finance, 7, 77-91.

B.2 Arbitrage

B.2.1 How to normalize prices ?

Let us first discuss alternative ways to normalize prices. For expository purpose, we consider the case of three goods with respective prices p_1, p_2, p_3. The decisions of the agents, that can be either consumers, firms or investors, depend on these prices up to a positive multiplicative factor. Thus, it is interesting to introduce a normalization to avoid this price multiplicity.

i) The most frequent normalization consists in choosing one of the good, good number 1, say, as a **numeraire**. Thus, the initial set of prices is replaced by 1, p_2/p_1, p_3/p_1. In economic reality, the money is generally used as the numeraire.

ii) However, this normalization is not the most appropriate in Finance, since it introduces an asymmetry between goods. Another possible normalization replaces the initial prices by:

$$q_1 = \frac{p_1}{p_1 + p_2 + p_3}, \quad q_2 = \frac{p_2}{p_1 + p_2 + p_3}, \quad q_3 = \frac{p_3}{p_1 + p_2 + p_3}. \tag{B.2.1}$$

This corresponds to the choice of the basket including one unit of each good as the numeraire. An advantage of the normalisation above is a possible interpretation of the new prices q_1, q_2, q_3 as a probability distribution, since $q_l \geq 0$, $\forall l$, and $\sum_{l=1}^{3} q_l = 1$.

This normalisation can also be applied to contingent assets. Let us consider an uncertain future with three states of nature w_1, w_2, w_3. An **Arrow-Debreu security** (or **digital option**) is an asset providing 1 money unit, if state w is realized, 0 money unit, otherwise. There exists in our example three Arrow-Debreu securities, with prices denoted by p_1, p_2, p_3, respectively. The basket including one unit of each Arrow-Debreu security provides one money unit with certainty. This is the z**ero-coupon bond**, whose price is

$B = \dfrac{1}{1+r_f}$, with r_f the riskfree interest rate. Thus, the prices of Arrow-Debreu securities can be normalized such that $p_j = Bq_j$, $j = 1, 2, 3$, where B is the price of the zero-coupon bond and q_j the elementary **risk-neutral probability**.

B.2.2 Absence of Arbitrage Opportunity (AAO)

The **Absence of Arbitrage Opportunity** assumes the impossibility to get a certain strictly positive future portfolio value for an initial nonpositive investment. It is also called assumption of **no arbitrage**, or of **no free lunch**. The AAO condition is automatically satisfied in an equilibrium model. Indeed, if a certain positive future value can be obtained from zero investment say, the investor will increase infinitely its investment size (the so-called **leverage effect**) implying an infinite demand of some assets, not compatible with the existence of an equilibrium.

There exist static and dynamic AAO condition. In the static case, the portfolio is crystallized at its initial allocation. In the dynamic case, the portfolio can be regularly updated without introducing or withdrawing cash at each updating (**self-financing condition**).

B.2.3 Pricing under dynamic AAO assumption

The no arbitrage condition implies strong restrictions between the asset prices. More precisely, let us consider a discrete time framework and assume an information I_t available to the investor when updating its portfolio at date t. Then the property below is providing a pricing formula.

Property B.1: *Let us consider a financial asset paying cash flows g_{t+h} at time $t+h$, where g_{t+h} depends on information I_{t+h}. Then, the price of this asset at time t can be written as:*

$$P(t, g) = \sum_{h=0}^{\infty} E_t[M_{t,t+h} g_{t+h}], \tag{B.2.2}$$

where $M_{t,t+h} = M_{t,t+1}M_{t+1,t+2} \dots M_{t+h-1,t+h}$*, with* $M_{t,t+1} \geq 0$ *depending on information* I_t*.*

The random variable $M_{t,t+1}$ is called the **short term stochastic discount factor** (sdf) and $M_{t,t+h}$ the sdf for term h. Thus, all asset prices are defined whenever the sequence of short term sdf is given. In general, the observed asset prices are not enough to characterize the underlying sdf. This is the **incompleteness** characteristic of the financial market.

The pricing formula (B.2.2) can be written in an alternative way.

i) Let us first consider a zero-coupon bond with time-to-maturity h. This bond provides a certain cash-flow equal to one at time $t+h$. Its price is equal to:

$$B(t,h) = E_t(M_{t,t+h}).$$

In particular $B(t,1) = E_t(M_{t,t+1}) \equiv \exp[-r(t,1)]$, where $r(t,1)$ is the continuously compounded riskfree short term rate.

Then, we get:

$$\begin{aligned}
& E_t(M_{t,t+h}g_{t+h}) \\
= \; & E_t\left\{\exp[-r(t,1)\dots - r(t+h-1,1)]\frac{M_{t,t+1}}{E_t(M_{t,t+1})}\dots\frac{M_{t+h-1,t+h}}{E_{t+h-1}(M_{t+h-1,t+h})}g_{t+h}\right\} \\
= \; & E_t^Q\left\{\exp[-r(t,1)\dots r(t+h-1,1)]g_{t+h}\right\},
\end{aligned}$$

where the **risk-neutral probability** Q admits (for time-to-maturity h) the density $\frac{M_{t,t+1}}{E_t(M_{t,t+1})}\dots\frac{M_{t+h-1,t+h}}{E_t(M_{t,+h-,t+h})}$ with respect to the initial probability distribution, called **historical distribution**, or **physical distribution**.

Corollary B.2: *Under the (dynamic) AAO condition, the asset price can be written as:*

$$P(t,g) = E_t^Q\left\{\exp[-r(t,1)\dots - r(t+h-1,1)]g_{t+h}\right\},$$

where $r(t,1)$ *is the short term riskfree rate and* Q *a risk-neutral distribution.*

The expression above corresponds to the second normalization discussed in Section B.2.1. The risk-neutral probability Q simply defines the normalized prices of appropriate Arrow-Debreu securities and is not unique in an incomplete market.

Further reading

Chamberlain, G., and M., Rothschild (1983): "Arbitrage, Factor Structure and Mean-Variance Analysis in Large Asset Markets", Econometrica, 52, 1281-1304.

Hansen, L., and S., Richard (1987): "The Role of Conditioning Information in Deducing Testable Restrictions Implied by Dynamic Asset Pricing Models", Econometrica, 55, 587-613.

Harrison, M., and D., Kreps (1979): "Martingales and Arbitrage in Multiperiod Securities Markets", Journal of Economic Theory, 20, 381-408.

Ross, S. (1976): "The Arbitrage Theory of Capital Asset Pricing", Journal of Economic Theory, 13, 341-360.

B.3 Risk Measures

The analysis of risky investments is based on quantities summarizing the risk, called **risk measures**. They can be used for descriptive purpose, but also for portfolio management, pricing, or definition of the required capital in a regulatory perspective.

The variance has long been the most successful measure of risk in Finance [see however Roy (1952)]. It is the basis of the mean-variance portfolio management [see Review B.1], and of the idea that the price of a risky asset is equal to its expected value plus a risk premium function of this variance. However, the variance (or the standard deviation), as a measure of risk, has some drawbacks. The surveys among professionals have shown that this risk measure was not so-well understood. Moreover, it is not appropriate to correctly capture the extreme risks, or the possible skew in the risk distribution. This has lead the regulatory authorities to choose the Value-at-Risk (VaR) as the new measure of risk in Basel I and II for banks, as well as in Solvency I and II for insurance companies. The aim of this review is to discuss some properties of the VaR, and its extensions, that are the Distortion Risk Measures (DRM).

Let us consider a random variable X, typically a Loss and Profit (L&P) variable, i.e. the opposite of a portfolio value, or the total liabilities in a balance. This variable has a distribution with a quantile function $q_\alpha(X)$ defined by:

$$P[X < q_\alpha(X)]) = \alpha.$$

Such a quantile function, called VaR in the regulation, characterizes the distribution of $L\&P$ variable X.

Definition B.1: *A risk measure $R(X)$ is a scalar function of the distribution of X, used to measure the risk.*

Of course not every function of this distribution is appropriate for measuring risk. Different conditions or axioms have been introduced in the literature to restrict the set of appropriate risk measures. We discuss below several of them.

i) The unit of a risk measure

Unit Axiom: *The risk measure has the same unit as variable X.*

This condition is important, if we want to use directly the risk measure as a level of reserve to hedge the risk, or as the cost (price) of this risk. For instance, the VaR and standard deviation satisfy the unit axiom, but not the variance. This shows also the importance of defining ex-ante the currency \$ or Euro, in which the risk measures are computed.

ii) Deterministic risk

Certainty Axiom: *If $X = c$ is known, then $R(X) = c$.*

This axiom shows that the search for a measure interpretable as a level of reserve, or as a price, has not only to account for the uncertainty of the value, but also for its "expected value". This condition is satisfied by the VaR, but not by the standard deviation.

iii) The homogeneity

Homogeneity Axiom: *We have $R(\lambda X) = \lambda R(X)$, $\forall \lambda > 0$.*

This condition is satisfied by both the VaR and the standard deviation. If $R(X)$ is seen as a price, the homogeneity axiom implies that the unitary price of an asset does not depend on the demanded quantity.

iv) Risk ordering

There exist two notions of risk ordering in the literature, both based on expected utility.

Definition B.2: *Let us consider two $L\&P$ variables X and Y. Variable X stochastically dominates variable Y at order 1 (resp. 2) if, and only if,*

$$E[U(-X)] \geq E[U(-Y)],$$

for any increasing (resp. increasing concave) function U.

The stochastic dominance at order j, for $j = 1, 2$, defines a preference ordering on the $L\&P$ variable X, or equivalently on the $P\&L$ variable $-X$.

Risk ordering Axiom: *We have $R(X) \leq R(Y)$, if X stochastically dominates Y.*

This condition on the risk measure is stronger with dominance at order 2 than with dominance at order 1.

v) Comonotonic risks

Two risks X and Y are comonotonous, if they are increasing functions of a same underlying risk Z: $X = a(Z)$, $Y = b(Z)$, say. Intuitively, they are increasing functions of a common risk factor.

Axiom of comonotonic risks: *$R(X+Y) = R(X) + R(Y)$, if X and Y are comonotonous.*

This axiom has been introduced to define reserve levels (or prices) in a way compatible with no arbitrage (see Review B.2). Typically, we have:

$$X - K = (X-K)^+ + [-(X-K)^-],$$

with $(X-K)^+ = \max(X-K, 0)$ and $(X-K)^- = \max(K-X, 0)$. Thus, the payoff is decomposed into the payoff of a European call with strike K and the payoff of a European put with the same strike. We expect to get:

$$R(X-K) = R[(X-K)^+] + R[-(X-K)^-],$$

to avoid a perfect arbitrage by means of reserves, whereas no arbitrage exists on the market.

The axiom on comonotonic risks is important and implies a first characterization of risk measures.

Property B.1: *The risk measures satisfying the axiom of comonotonic risks, the certainty axiom and the compatibility with first-order stochastic dominance can be written as:*

$$R(X) = \int_0^1 q_\alpha(X) dH(\alpha),$$

where H is the cumulative distribution function of a probability distribution on $[0,1]$.

Measure H is called a **Distortion Measure** (DM) and R a **Distortion Risk Measure** (DRM). A distortion risk measure is simply a weighted combination of VaR at several quantile levels. The VaR at level α is itself a DRM by choosing the point mass at α as distortion measure.

When the distortion measure is the uniform distribution on $(\alpha, 1)$, the DRM reduces to the **Expected Shortfall** (ES) at level α, given by:

$$ES_\alpha(X) = E[X|X > q_\alpha(X)] = \frac{1}{1-\alpha}\int_\alpha^1 q_u(X)du. \tag{1.1}$$

Corollary B.2: *A DRM is compatible with second-order stochastic dominance if and only if the distortion cdf H is convex.*

The Expected Shortfall satisfies this condition, but not the VaR.

vi) Subadditivity

Subadditivity Axiom: *For any risks X and Y, we have:*

$$R(X+Y) \leq R(X) + R(Y)$$

The DRMs with convex distortion measure satisfy this axiom, but not the VaR (even if we observe $VaR(X+Y) \leq VaR(X) + VaR(Y)$ for the portfolios risks encountered in practice).

The subadditivity condition is a source of debate among academics and practitioners, especially when it is used as a crude tool for fixing regulatory reserves. Let us assume that a regulator demands to each bank $i = 1, \ldots, n$ to fix its required capital at $R(X_i)$, where R is a subadditive risk measure (for instance the expected shortfall at 95%). Then,

$$\sum_{i=1}^n R(X_i) \geq R(\sum_{i=1}^n X_i).$$

At a first sight, the regulator oversizes the capital required to hedge the global risk $X = \sum_{i=1}^{n} X_i$, that is, seems to follows a prudential approach. However, with such a principle, we also have:

$$R(X_1 + X_2) \leq R(X_1) + R(X_2),$$

which is a strong incentive for banks 1 and 2 to merge to diminuish the level of required capital. Thus, this a priori prudential approach can have spurious consequences.

The risk measures satisfying the certainty axiom, the homogeneity axiom, the subadditivity axiom and compatible with second-order stochastic dominance are called **coherent risk measures**. The coherent risk measures can be written as:

$$R(X) = \sup_{H \in \mathcal{H}} DRM_H(X),$$

that is as a supremom of a set of convex DRM risk measures.

Further reading

Acerbi, C., and D., Tasche (2002): "Expected Shortfall : a Natural Coherent Alternative to Value-at-Risk", Economic Notes, 31, 379-388.

Artzner, P., Delbaen, F., Eber, J., and D., Heath (1997): "Thinking Coherently", Risk, 10,

Gourieroux, C., and A., Monfort (2011): “Allocating Systematic and Unsystematic Risks in a Regulatory Perspective”, CREST DP.

Kusuoka, S. (2001): "On Law Invariant Coherent Risk Measures", Advances in Mathematical Economics, 3, 83-95.

Roy, A. (1952): "Safety First and the Holding of Assets", Econometrica, 20, 431-449.

Wang, S. (1996): "Premium Calculation by Transforming the Layer Premium Density", ASTIN Bulletin, 26, 71-92.

Wang, S., Young, V., and H., Panjer (1997): "Axiomatic Characterization of Insurance Prices", Insurance: Mathematics and Economics, 21, 173-183.