Microinformation, Nonlinear Filtering and Granularity

Patrick Gagliardini *

Christian Gouriéroux †

Alain Monfort ‡

First Version: November 2009

This Version: July 2010 §

*University of Lugano and Swiss Finance Institute.

†CREST, CEPREMAP (Paris) and University of Toronto.

‡CREST, Banque de France and Maastricht University.

§Acknowledgements: This paper has been prepared for the JFEC Lecture given by the second author at the SoFiE conference 2010 in Melbourne. We thank D. Bates and N. Shephard for useful comments. The first author gratefully acknowledges financial support of the Swiss National Science Foundation through the NCCR FINRISK network. The second author gratefully acknowledges financial support of NSERC Canada and the chair AXA/Risk Foundation: "Large Risks in Insurance".

Microinformation, Nonlinear Filtering and Granularity

Abstract

The recursive prediction and filtering formulas of the Kalman filter are difficult to implement in nonlinear state space models. For Gaussian linear state space models, or for models with qualitative state variables, the recursive formulas of the filter require the updating of a finite number of summary statistics. However, in the general framework a function has to be updated, which makes the approach computationally cumbersome. The aim of this paper is to consider the situation of a large number n of individual measurements, the so-called microinformation, and to take advantage of the large cross-sectional size to get closed-form prediction and filtering formulas at order $1/n$. The state variables have a macro-factor interpretation. The results are applied to the maximum likelihood estimation of a macro-parameter, and to the computation of a granularity adjusted Value-at-Risk (VaR) for large portfolios. The methodology of granularity adjustment for VaR is illustrated by an application of the Value of the Firm model [Merton (1974)] to both default and loss given default.

Keywords: Kalman Filter, Nonlinear State Space, Granularity, Repeated Observations, Value-at-Risk, Credit Risk, Loss Given Default, Basel 2.
JEL classification: G12, C23.

1 Introduction

Let us consider a nonlinear state space model with observations y_t, $t = 1, \cdots, T$, and underlying latent state variables F_t. We denote by $\mathbf{Y}_t$ (resp. $\mathbf{F}_t$) the information included in the current and past values of variable y (resp. F). The model is defined by (i) the state equation, which specifies the conditional probability distribution function (pdf) of F_t given $\mathbf{F}_{t-1}$, $\mathbf{Y}_{t-1}$ as $g(f_t|f_{t-1})$, say; (ii) the measurement equation, which specifies the conditional pdf of y_t given $\mathbf{F}_t$, $\mathbf{Y}_{t-1}$ as $h(y_t|f_t)$, say. Thus, the state variable is assumed to follow an autonomous Markov process of order 1, and the distribution of the observed variable depends on the information through the current factor value only [1]. In such a nonlinear state space model, the joint pdf of the observations (given some initial condition) is:

$$\int \prod_{t=1}^{T} [h(y_t|f_t)g(f_t|f_{t-1})] \prod_{t=1}^{T} df_t, \tag{1.1}$$

and involves a multiple integral with dimension equal to sample size T times the dimension of the underlying factor.

The nonlinear Kalman filter proposes a recursive computation of well-chosen conditional distributions. The filtering density provides the conditional pdf $p(f_t|\mathbf{Y}_t)$ of factor F_t given $\mathbf{Y}_t$. The predictive density provides the conditional pdf of y_{t+1} given $\mathbf{Y}_t$, denoted $p(\tilde{y}_{t+1}|\mathbf{Y}_t)$, where $\tilde{y}_{t+1}$ indicates a generic argument of variable y_{t+1}. Then the joint pdf of the sample observations is deduced by multiplying the successive predictive densities, evaluated at the observed values $\tilde{y}_{t+1} = y_{t+1}$ for $t = 0, 1, \cdots, T-1$.

Let us recall some recursions involved in the nonlinear Kalman filter. We have for instance:

$$\begin{aligned} p(\tilde{y}_{t+1}|\mathbf{Y}_t) &= E\left[p(\tilde{y}_{t+1}|\mathbf{F}_t, \mathbf{Y}_t)|\mathbf{Y}_t\right] \\ &= E\left[\int h(\tilde{y}_{t+1}|f_{t+1})g(f_{t+1}|F_t)df_{t+1}|\mathbf{Y}_t\right] \\ &= E\left[\Psi(\tilde{y}_{t+1}, F_t)|\mathbf{Y}_t\right], \end{aligned}$$

where:

$$\Psi(\tilde{y}_{t+1}, f_t) = \int h(\tilde{y}_{t+1}|f_{t+1})g(f_{t+1}|f_t)df_{t+1}. \tag{1.2}$$

[1] This model is sometimes called Hidden Markov Model (HMM) in the literature [see e.g. Cappé, Moulines, Rydén (2005) for a review on inference in HMM]. It is extended in Section 2 to allow for the effect of exogenous regressors and lagged observations in the measurement equation.

Thus, we get the updating of the predictive distribution from the filtering distribution:

$$p(\tilde{y}_{t+1}|\mathbf{Y}_t) = \int \Psi(\tilde{y}_{t+1}, f_t) p(f_t|\mathbf{Y}_t) df_t. \tag{1.3}$$

The integrals in (1.2) and (1.3) often have a small dimension and could be easily computed numerically. However, this type of updating formula is difficult to implement in the general framework, since it requires as input the function $p(\cdot|\mathbf{Y}_t)$. Hence, it is necessary to temporarily store this function at each recursion [2]. Three special cases are known, in which the nonlinear Kalman filter is simplified, because only a finite number of scalars have to be updated. These are the Gaussian linear state space model, initially considered by Kalman [Kalman (1960), Kalman and Bucy (1961)], the model with qualitative factor, at the core of the Kitagawa filter [Kitagawa (1987), (1996), Hamilton (1989)], and state space models with finite-dimensional dependence [Gouriéroux, Jasiak (2002)].

This paper introduces another framework in which the nonlinear Kalman filter can be (approximately) solved in closed-form. Specifically, we consider a large number n of individual measurements $y_t = (y_{1,t}, \cdots, y_{n,t})'$, and exploit the cross-sectional dimension to approximate the nonlinear Kalman filter at order $1/n$.

The model and the approximate prediction and filtering formulas are given in Section 2. The special case of measurement model in an exponential family is discussed in Section 3. In Section 4, we consider the estimation of a macro-parameter in a model with Gaussian factor. For this estimation problem, we show that the approximate nonlinear Kalman filter to compute the joint distribution of the observations is equivalent to a standard Kalman filter applied to an approximate linear state space model. An application to the computation of the Value-at-Risk (VaR) for a large homogeneous portfolio is discussed in Section 5. In Section 6 the above methodology is applied to Merton's model for credit risk [Merton (1974)], when both default and Loss Given Default (LGD) are taken into account. Section 7 concludes. Proofs are gathered in appendices. For simplicity, we focus on the most common case of a single factor. The results can be generalized to multiple factors, but the derivations are notationally cumbersome at some steps.

[2] This is usually considered as an issue of numerical approximation. Simulation-based approaches include sequential Monte-Carlo methods such as particle filtering, where the filtering distribution is approximated by "particles" with discrete probability mass [see e.g. Pitt, Shephard (2001), and Cappé, Moulines, Rydén (2005), Chapter 7]. However, it may not be easy to control the associated approximation error.

2 Approximate Prediction and Filtering

2.1 The Nonlinear State Space Model

The observations are endogenous individual variables $y_{i,t}$, for $i = 1, \cdots, n$, $t = 1, \cdots, T$, and exogenous variables x_i, for $i = 1, \cdots, n$. The latter variables are indexed by individual i only and correspond to time invariant individual characteristics [3]. The state variables F_t are indexed by time t only. They are unobservable and can be interpreted as macro-factors (or as systematic risk factors in financial applications). We denote by $y_t = (y_{1,t}, \cdots, y_{n,t})'$ [resp. $X = (x_1', \cdots, x_n')'$] the set of cross-sectional observations on y (resp. on x).

As usual, the nonlinear state space model is defined by measurement and state equations, given below in terms of conditional distributions.

State equation: *The conditional distribution of F_t given $\mathbf{F}_{t-1}$, $\mathbf{Y}_{t-1}$, X depends on F_{t-1} only, is time-invariant, and admits a pdf $g(f_t|f_{t-1})$, $t = 1, \cdots, T$.*

Measurement equations: *Conditionally on the information set $\mathbf{F}_t$, $\mathbf{Y}_{t-1}$, X, the individual endogenous variables $y_{i,t}$, $i = 1, \cdots, n$, are independent. The distribution of $y_{i,t}$ given $\mathbf{F}_t$, $\mathbf{Y}_{t-1}$, X depends on F_t, $y_{i,t-1}$ and x_i only, is time-invariant and admits the pdf:*

$$h(y_{i,t}|f_t, y_{i,t-1}, x_i) \equiv h_{i,t}(y_{i,t}|f_t), \ i = 1, \cdots, n, \ t = 1, \cdots, T.$$

This nonlinear state space model allows for exogenous variables in the measurement equations, introducing observable heterogeneity across individuals. It also allows for both a micro-dynamics by means of the individual lags in the measurement equations, and a macro-dynamics by means of the unobservable factors. The model includes as a special case models with repeated observations when $h_{i,t}(y_{i,t}|f_t) = h(y_{i,t}|f_t)$.

[3] The approximate filtering and predictive distributions at horizon 1 derived in the paper are also valid when observable macro-variables z_t, say, are introduced in the state equation, and possibly time dependent individual exogenous variables $z_{i,t}$, say, in the measurement equations. However, as usual in state space models, the filtering and predictive distributions at horizon strictly larger than 1 require the specification of the dynamics of the variables $z_{i,t}$ and z_t.

The value of the unobservable factor F_t can be approximated by the cross-sectional maximum likelihood (CSML) estimator defined by:

$$\hat{f}_{n,t} = \arg\max_{f_t} \sum_{i=1}^{n} \log h_{i,t}(y_{i,t}|f_t). \tag{2.1}$$

The terminology CSML is convenient but a bit abusive since, if the micro-density $h_{i,t}(y_{i,t}|f_t;\beta)$ depends on an unknown micro-parameter β, the CSML estimator $\hat{f}_{n,t}(\beta)$ also depends on β. In some sense we are concentrating the micro log-likelihood function with respect to f_t considered as a "nuisance" parameter. If parameter β is known, $\hat{f}_{n,t}(\beta)$ provides an approximation of factor f_t, which is consistent if the cross-sectional size n tends to infinity. However, it is not the most accurate one, since it does not take into account the lagged observations of y and the factor dynamics. We will see later on that the cross-sectional approximation of the factor plays a crucial role in the derivation of the prediction and filtering formulas.

Other cross-sectional summary statistics will also be useful. Let us introduce the notation:

$$K_{n,t}^{(p)} = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial^p \log h_{i,t}(y_{i,t}|\hat{f}_{n,t})}{\partial f_t^p}, \; p = 2, 3, 4. \tag{2.2}$$

The quantity:

$$I_{n,t} = -K_{n,t}^{(2)}, \tag{2.3}$$

measures the accuracy of $\hat{f}_{n,t}$ as an approximation of f_t (with known β); the quantity $K_{n,t}^{(3)}$ is involved in the bias at order $1/n$ of estimator $\hat{f}_{n,t}$. Under appropriate stationarity assumptions, the quantities $K_{n,t}^{(p)}$ are $O_P(1)$, when n tends to infinity.

2.2 Approximate Filtering Formula

An approximation of the filtering distribution for factor F_t is derived by means of the Laplace method [see e.g. Jensen (1995)]. The form of the approximation is given in the next Proposition 1 (see Appendix 1 for the proof). This result extends the approximate filtering distribution derived in Gagliardini and Gouriéroux (2009) to a model with micro-dynamics and exogenous variables.

PROPOSITION 1: *At order $1/n$, the conditional distribution of F_t given $\mathbf{Y}_t$, $\mathbf{F}_{t-1}$, X is equal to the conditional distribution of F_t given $\mathbf{Y}_t$, X only, i.e. to the filtering distribution. This distribu-*

tion is Gaussian and is given by:

$$N\left(\hat{f}_{n,t} + \frac{1}{n}\left[I_{n,t}^{-1}\frac{\partial \log g}{\partial f_t}(\hat{f}_{n,t}|\hat{f}_{n,t-1}) + \frac{1}{2}I_{n,t}^{-2}K_{n,t}^{(3)}\right], \frac{1}{n}I_{n,t}^{-1}\right).$$

At order $1/n$, the filtering distribution of F_t differs from a point mass at the CSML estimate $\hat{f}_{n,t}$. By extending the notion of granularity introduced by Gordy (2003) in the context of portfolio VaR computation, we call this distribution the granularity adjusted (GA) filtering distribution. The variance of the GA filtering distribution shrinks to zero at rate $1/n$ and the mean of the filtering distribution differs from $\hat{f}_{n,t}$ by a term of order $1/n$. The granularity adjustment involves the four summary statistics $\hat{f}_{n,t}$, $\hat{f}_{n,t-1}$, $I_{n,t}$, $K_{n,t}^{(3)}$. The dynamics of the latent factor impacts the filtering distribution through the partial derivative of the log transition pdf $\frac{\partial \log g}{\partial f_t}(\hat{f}_{n,t}|\hat{f}_{n,t-1})$. Finally, conditionally on $\mathbf{Y}_t$ and X, the current and the lagged factors F_t and $\mathbf{F}_{t-1}$ are independent at order $1/n$.

The Gaussian approximate filtering distribution in Proposition 1 shares some common features with the approximations considered in the literature on robust Kalman filtering [see e.g. Masreliez (1975)]. However, it differs in several respects. First, in robust filtering the conditional distribution of F_{t+1} given $\mathbf{Y}_t$ is assumed to be close to a Gaussian distribution, whereas in our framework it is the conditional distribution of F_t given $\mathbf{Y}_t$ which is almost Gaussian [4]. Second, in robust filtering the errors of the analytical approximations are typically unknown [5], while in our approach the Gaussian approximation has been derived theoretically together with its approximation error due to the information contained in the cross-sectional observations. Third, the robust filtering literature mostly focuses on linear measurement and state equations with non-Gaussian innovations [6], while our model fully allows for nonlinearities in both the measurement and state equations. Finally, the approximation in Proposition 1 is not recursive, but in closed form.

[4]See Bates (2009), p. 25, for approximations written on the same conditional distribution as ours. These approximations are used in the numerical implementation of an algorithm that updates the Laplace transform of the filtering distribution when the joint dynamics of observations and latent states is affine.

[5]Except in the special model of contamination considered in Schick, Mitter (1994).

[6]Except for instance Cipra and Rubio (1991), who take into account a nonlinear measurement equation with additive non-Gaussian innovations.

2.3 Approximate Prediction Formula

The approximate filtering formula in Proposition 1 can be used to derive the prediction formula at order $1/n$, that is, the conditional distribution of y_{t+1} given $\mathbf{Y}_t$, X. More precisely, we have by the law of iterated expectation:

$$p\left(\tilde{y}_{t+1}|\mathbf{Y}_t, X\right) = E\left[p\left(\tilde{y}_{t+1}|\mathbf{Y}_t, \mathbf{F}_t, X\right)|\mathbf{Y}_t, X\right] = E\left[\Psi(\tilde{y}_{t+1}|y_t, F_t, X)|\mathbf{Y}_t, X\right],$$

where $\Psi(\tilde{y}_{t+1}|y_t, F_t, X) = p\left(\tilde{y}_{t+1}|\mathbf{Y}_t, \mathbf{F}_t, X\right)$ depends on the past through y_t and F_t only because of the assumptions on the state and measurement equations. Thus, the derivation of the predictive distribution can be performed in two steps. We first derive an approximation at order $1/n$ of the conditional distribution of y_{t+1} given F_t, y_t and X; then, F_t is integrated out using its conditional pdf given $\mathbf{Y}_t$ and X in Proposition 1.

The conditional pdf of y_{t+1} given y_t, F_t, X is:

$$\Psi(\tilde{y}_{t+1}|y_t, f_t, X) = \int \prod_{i=1}^{n} h_{i,t+1}(\tilde{y}_{i,t+1}|f_{t+1}) g(f_{t+1}|f_t) df_{t+1}.$$

This pdf can be written as:

$$\Psi(\tilde{y}_{t+1}|y_t, f_t, X) = \int \exp\left[\sum_{i=1}^{n} \log h_{i,t+1}(\tilde{y}_{i,t+1}|f_{t+1}) + \log g(f_{t+1}|f_t)\right] df_{t+1}. \tag{2.4}$$

The integrand can be expanded around the cross-sectional approximation $\tilde{f}_{n,t+1}$ to get the result below (see Appendix 2), where $\tilde{f}_{n,t+1}$ is the CSML estimator of f_{t+1} based on $\tilde{y}_{t+1}$, y_t, X. Similarly, we denote by $\tilde{K}^{(p)}_{n,t+1}$, $\tilde{I}_{n,t+1}$ the summary statistics with y_{t+1} replaced by the generic argument $\tilde{y}_{t+1}$.

PROPOSITION 2: *At order $1/n$, the conditional pdf of y_{t+1} given y_t, F_t, X is equal to:*

$$\begin{aligned}\Psi(\tilde{y}_{t+1}|y_t, f_t, X) \quad = \quad & \sqrt{\frac{2\pi}{n\tilde{I}_{n,t+1}}} \prod_{i=1}^{n} h_{i,t+1}(\tilde{y}_{i,t+1}|\tilde{f}_{n,t+1}) g(\tilde{f}_{n,t+1}|f_t) \\ & \cdot \exp\left\{\frac{1}{n}\left[\frac{1}{8}\tilde{K}^{(4)}_{n,t+1}\tilde{I}^{-2}_{n,t+1} + \frac{1}{2}\tilde{I}^{-1}_{n,t+1}\left(\frac{\partial^2 \log g(\tilde{f}_{n,t+1}|f_t)}{\partial f^2_{t+1}} + \left(\frac{\partial \log g(\tilde{f}_{n,t+1}|f_t)}{\partial f_{t+1}}\right)^2\right)\right.\right. \\ & \left.\left. + \frac{1}{2}\tilde{I}^{-2}_{n,t+1}\tilde{K}^{(3)}_{n,t+1}\frac{\partial \log g(\tilde{f}_{n,t+1}|f_t)}{\partial f_{t+1}} + \frac{5}{24}\left[\tilde{K}^{(3)}_{n,t+1}\right]^2 \tilde{I}^{-3}_{n,t+1}\right] + o(1/n)\right\}.\end{aligned}$$

The normalization factor $\sqrt{2\pi/n}$ ensures that the integral of $\Psi(\tilde{y}_{t+1}|y_t, f_t, X)$ w.r.t. $\tilde{y}_{t+1}$ is equal to 1 at order $o(1/n)$. Alternatively, we could impose the exact validity of the unit mass restriction by normalizing the approximate density by its numerical integral.

Then, the expression of Proposition 2 can be integrated w.r.t. the approximate Gaussian filtering distribution of F_t given in Proposition 1 in order to get the predictive pdf. We obtain the following result:

PROPOSITION 3: *At order $1/n$, the predictive pdf of y_{t+1} given $\boldsymbol{Y}_t$, X is equal to:*

$$\begin{aligned} p(\tilde{y}_{t+1}|\boldsymbol{Y}_t, X) \;=\; & \sqrt{\frac{2\pi}{n\tilde{I}_{n,t+1}}} \prod_{i=1}^{n} h_{i,t+1}\left(\tilde{y}_{i,t+1}|\tilde{f}_{n,t+1}\right) g\left(\tilde{f}_{n,t+1}|\hat{f}_{n,t}\right) \\ & \cdot \exp\left\{ \frac{1}{n}\left[\frac{1}{8}\tilde{K}^{(4)}_{n,t+1}\tilde{I}^{-2}_{n,t+1} + \frac{5}{24}\left[\tilde{K}^{(3)}_{n,t+1}\right]^2 \tilde{I}^{-3}_{n,t+1} \right.\right. \\ & + \frac{1}{2}\tilde{I}^{-1}_{n,t+1}\left(\frac{\partial^2 \log g\left(\tilde{f}_{n,t+1}|\hat{f}_{n,t}\right)}{\partial f^2_{t+1}} + \left(\frac{\partial \log g\left(\tilde{f}_{n,t+1}|\hat{f}_{n,t}\right)}{\partial f_{t+1}} \right)^2 \right) \\ & + \frac{1}{2}I^{-1}_{n,t}\left(\frac{\partial^2 \log g\left(\tilde{f}_{n,t+1}|\hat{f}_{n,t}\right)}{\partial f^2_{t}} + \left(\frac{\partial \log g\left(\tilde{f}_{n,t+1}|\hat{f}_{n,t}\right)}{\partial f_{t}} \right)^2 \right) \\ & + I^{-1}_{n,t}\frac{\partial \log g\left(\tilde{f}_{n,t+1}|\hat{f}_{n,t}\right)}{\partial f_t}\frac{\partial \log g\left(\hat{f}_{n,t}|\hat{f}_{n,t-1}\right)}{\partial f_t} \\ & \left.\left. + \frac{1}{2}\tilde{I}^{-2}_{n,t+1}\tilde{K}^{(3)}_{n,t+1}\frac{\partial \log g\left(\tilde{f}_{n,t+1}|\hat{f}_{n,t}\right)}{\partial f_{t+1}} + \frac{1}{2}I^{-2}_{n,t}K^{(3)}_{n,t}\frac{\partial \log g\left(\tilde{f}_{n,t+1}|\hat{f}_{n,t}\right)}{\partial f_t} \right] + o(1/n) \right\}. \end{aligned}$$

We get a closed form expression for the predictive density. It depends on summary statistics $\tilde{f}_{n,t+1}$, $\tilde{I}_{n,t+1}$, $\tilde{K}^{(3)}_{n,t+1}$, $\tilde{K}^{(4)}_{n,t+1}$, $\hat{f}_{n,t}$, $\hat{f}_{n,t-1}$, $I_{n,t}$, $K^{(3)}_{n,t}$, some of them being functions of the selected argument $\tilde{y}_{t+1}$. The formula in Proposition 3 is simplified when the argument of interest $\tilde{y}_{t+1} = y_{t+1}$ corresponds to the observations, as for deriving the joint density function of the sample. Indeed, in this case, we have $\tilde{f}_{n,t+1} = \hat{f}_{n,t+1}$, $\tilde{I}_{n,t+1} = I_{n,t+1}$ and $\tilde{K}^{(p)}_{n,t+1} = K^{(p)}_{n,t+1}$. In particular, we see that process (y_t) is a Markov process of order 2, up to $o(1/n)$.

3 Exponential Micro-model

The expressions for the filtering and prediction distributions in Section 2 capture the non-Gaussianity of both the micro- and macro-dynamics. This effect is illustrated in this section for a model with exponential micro-density.

3.1 The Model

Let us assume that the conditional micro-density can be written as:

$$h_{i,t}(y_{i,t}|f_t) = \exp\left[a(y_{i,t})f_t + b(y_{i,t}) + c(f_t)\right]. \tag{3.1}$$

This is an exponential family in which the factor value is the canonical parameter. We have the following property (see Appendix 4 for the proof):

PROPOSITION 4: *For an exponential micro-model with canonical factor F_t, we have:*

$$K_{n,t}^{(p)} = \frac{d^p c(\hat{f}_{n,t})}{df_t^p},\ p \geq 2.$$

Moreover:

$$\begin{aligned} \frac{d^2 c(f_t)}{df_t^2} &= -V[a(y_{i,t})|F_t = f_t], \\ \left[-\frac{d^2 c(f_t)}{df_t^2}\right]^{-3/2} \frac{d^3 c(f_t)}{df_t^3} &= -\text{Skewness}[a(y_{i,t})|F_t = f_t], \\ \left[-\frac{d^2 c(f_t)}{df_t^2}\right]^{-2} \frac{d^4 c(f_t)}{df_t^4} &= -\text{Excess Kurtosis}[a(y_{i,t})|F_t = f_t]. \end{aligned}$$

Therefore, the adjustment at order $1/n$ in the filtering distribution (Proposition 1) involving the third-order derivative of the micro-density contains among other statistics the measure of conditional skewness $I_{n,t}^{-3/2}K_{n,t}^{(3)}$. Similarly, the adjustments in the predictive distribution (Proposition 3) involve both conditional skewness and excess kurtosis measures, through statistics $I_{n,t}^{-3/2}K_{n,t}^{(3)}$, $\tilde{I}_{n,t}^{-3/2}\tilde{K}_{n,t}^{(3)}$ and $\tilde{I}_{n,t}^{-2}\tilde{K}_{n,t}^{(4)}$. Skewness and excess kurtosis summarize the properties of the conditional distribution of the transform $a(y_{i,t})$ of the individual observation given the factor value, that are involved in the adjustments at order $1/n$.

3.2 Examples

We provide in Table 1 the canonical parameterization and the main summary statistics for standard exponential families. For some of them (e.g., the Bernoulli family), the canonical parameterization does not coincide with the usual parameterization. From function $c(f)$ and the cross-sectional ML estimator of the factor value $\hat{f}_{n,t}$, we can deduce the expressions of the statistics $K_{n,t}^{(p)}$.

Example 1: Gaussian family with factor in mean

For a linear Gaussian state space model, the measurements are such that $y_{1,t}, \cdots, y_{n,t} \sim IIN(f_t, 1)$ conditional on $F_t = f_t$, where the canonical factor value f_t is the conditional mean, and the conditional variance is constant, equal to 1, say. The CSML estimator of the factor value is $\hat{f}_{n,t} = \frac{1}{n}\sum_{i=1}^{n} y_{i,t}$, that is the cross-sectional average of the observations at date t. The statistics $K_{n,t}^{(p)}$ are such that $I_{n,t} = -K_{n,t}^{(2)} = 1$ and $K_{n,t}^{(p)} = 0$ for $p > 2$.

Example 2: Bernoulli family with stochastic probability

For qualitative observations in the Bernoulli family, we have $y_{1,t}, \cdots, y_{n,t} \sim i.i.\mathcal{B}(1, p_t)$ conditionally on $F_t = f_t$, where the canonical factor value f_t is related with the conditional probability p_t by $f_t = \log[p_t/(1-p_t)]$. The CSML estimator of the factor value is $\hat{f}_{n,t} = \log[\bar{y}_{n,t}/(1-\bar{y}_{n,t})]$, where $\bar{y}_{n,t} = \frac{1}{n}\sum_{i=1}^{n} y_{i,t}$ is the cross-sectional frequency. The statistics $K_{n,t}^{(p)}$ are such that $I_{n,t} = -K_{n,t}^{(2)} = \bar{y}_{n,t}(1-\bar{y}_{n,t})$, $K_{n,t}^{(3)} = -\bar{y}_{n,t}(1-\bar{y}_{n,t})(1-2\bar{y}_{n,t})$ and $K_{n,t}^{(4)} = -\bar{y}_{n,t}(1-\bar{y}_{n,t})(1-6\bar{y}_{n,t}+2\bar{y}_{n,t}^2)$.

4 Gaussian Factor and ML Estimation of Macro-parameters

In a nonlinear state space model, the unobservable factor is defined up to a one-to-one (nonlinear) transformation. We have seen in Section 3, that the choice of a canonical factor is useful to interpret asymptotic adjustments in the filtering and prediction distributions. In practice, however, it is also useful to select factors with Gaussian autoregressive dynamics; this requires factors which can take real negative and positive values.

In Example 2 with the Bernoulli family the canonical factor $f = \log[p/(1-p)]$ admits real values, but in other cases the canonical factor is constrained. For instance, in the exponential family in Table 1, the canonical factor $f = \lambda$ is positive, as well as in the Gaussian model with volatility factor.

In this Section we consider a model with stationary Gaussian autoregressive factor:

$$F_t = \mu + \gamma(F_{t-1} - \mu) + \eta\sqrt{1-\gamma^2}\varepsilon_t, \tag{4.1}$$

where the innovations are $\varepsilon_t \sim IIN(0,1)$ and the autoregressive coefficient γ is such that $|\gamma| < 1$. The stationary distribution of F_t is Gaussian with mean μ and variance η^2. The transition pdf is:

$$g(f_t|f_{t-1};\theta) = \frac{1}{\sqrt{2\pi\eta^2(1-\gamma^2)}} \exp\left\{-\frac{[f_t - \mu - \gamma(f_{t-1}-\mu)]^2}{2\eta^2(1-\gamma^2)}\right\}, \tag{4.2}$$

where the macro-parameter $\theta = (\mu, \gamma, \eta^2)'$ is unknown. In this section we also assume that the micro-density $h(y_{i,t}|f_t)$, say, is completely known, and we consider maximum likelihood estimation of parameter θ. [7]

4.1 Approximate Log-likelihood Function

The exact log-likelihood function (conditional on the initial observation) can be written as:

$$\mathcal{L}_{nT}(\theta) = \mathcal{L}_{nT}^{\mathrm{GA}}(\theta) + o_p(1/n), \tag{4.3}$$

where the granularity adjusted (GA) likelihood function is given by:

$$\mathcal{L}_{nT}^{\mathrm{GA}}(\theta) = \frac{1}{T}\sum_{t=1}^{T} \log p^{GA}(y_t|\mathbf{Y}_{t-1}, X;\theta),$$

and $p^{GA}(y_t|\mathbf{Y}_{t-1}, X;\theta)$ is the approximate predictive density at order $1/n$ given in Proposition 3 and evaluated at the sample observations $\tilde{y}_t = y_t$. Therefore, instead of considering the unfeasible maximum likelihood estimator:

$$\hat{\theta}_{nT} = \arg\max_{\theta} \mathcal{L}_{nT}(\theta),$$

we can consider the approximation obtained by maximizing the GA log-likelihood function:

$$\hat{\theta}_{nT}^{\mathrm{GA}} = \arg\max_{\theta} \mathcal{L}_{nT}^{\mathrm{GA}}(\theta). \tag{4.4}$$

[7] See Gagliardini and Gouriéroux (2010a) for the general case where the micro-density involves the lagged endogenous variable $y_{i,t-1}$ and an unknown parameter β, that is, $h(y_{i,t}|y_{i,t-1}, f_t;\beta)$.

The granularity adjusted estimator differs from the unfeasible maximum likelihood estimator by a term negligible at order $1/n$ [see Gagliardini, Gouriéroux (2010a)]:

$$\hat{\theta}_{nT}^{\text{GA}} - \hat{\theta}_{nT} = o_p(1/n).$$

Let us now focus on the granularity adjusted log-likelihood function. We can write:

$$\mathcal{L}_{nT}^{\text{GA}}(\theta) = \mathcal{L}_{0,nT} + \frac{1}{T}\sum_{t=1}^{T} \log p_1(y_t|\mathbf{Y}_{t-1}, X; \theta), \tag{4.5}$$

where $\mathcal{L}_{0,nT}$ is independent of θ and:

$$\begin{aligned}
\log p_1(y_t|\mathbf{Y}_{t-1}, X; \theta) \;=\; & \log g\left(\hat{f}_{n,t}|\hat{f}_{n,t-1};\theta\right) \\
& + \frac{1}{2n} I_{n,t}^{-1} \left(\frac{\partial^2 \log g\left(\hat{f}_{n,t}|\hat{f}_{n,t-1};\theta\right)}{\partial f_t^2} + \left(\frac{\partial \log g\left(\hat{f}_{n,t}|\hat{f}_{n,t-1};\theta\right)}{\partial f_t} \right)^2 \right) \\
& + \frac{1}{2n} I_{n,t-1}^{-1} \left(\frac{\partial^2 \log g\left(\hat{f}_{n,t}|\hat{f}_{n,t-1};\theta\right)}{\partial f_{t-1}^2} + \left(\frac{\partial \log g\left(\hat{f}_{n,t}|\hat{f}_{n,t-1};\theta\right)}{\partial f_{t-1}} \right)^2 \right) \\
& + \frac{1}{n} I_{n,t-1}^{-1} \frac{\partial \log g\left(\hat{f}_{n,t}|\hat{f}_{n,t-1};\theta\right)}{\partial f_{t-1}} \frac{\partial \log g\left(\hat{f}_{n,t-1}|\hat{f}_{n,t-2};\theta\right)}{\partial f_{t-1}} \\
& + \frac{1}{2n} I_{n,t}^{-2} K_{n,t}^{(3)} \frac{\partial \log g\left(\hat{f}_{n,t}|\hat{f}_{n,t-1};\theta\right)}{\partial f_t} + \frac{1}{2} I_{n,t-1}^{-2} K_{n,t-1}^{(3)} \frac{\partial \log g\left(\hat{f}_{n,t}|\hat{f}_{n,t-1};\theta\right)}{\partial f_{t-1}}.
\end{aligned} \tag{4.6}$$

When the transition p.d.f. of the factor corresponds to the Gaussian autoregressive model (4.2), the log-density $\log g\left(\hat{f}_{n,t}|\hat{f}_{n,t-1};\theta\right)$ and its partial derivatives in the RHS of (4.6) are polynomials in $\hat{f}_{n,t} - \mu - \gamma(\hat{f}_{n,t-1} - \mu)$ and $\hat{f}_{n,t-1} - \mu - \gamma(\hat{f}_{n,t-2} - \mu)$ of degree less or equal to 2. This explains why the GA log-likelihood function is equivalent to the logarithm of a Gaussian pdf for $\hat{f}_{n,t} - \mu - \gamma(\hat{f}_{n,t-1} - \mu)$, $t = 1, \cdots, T$, with granularity adjustments for the mean and the variance-covariance structure at order $1/n$ (see Appendix 5). We get the next result.

PROPOSITION 5: *In a model with Gaussian autoregressive factor and macro-parameter θ only, a granularity adjusted maximum likelihood estimator of θ can be obtained by maximizing the likelihood function of the (time-inhomogeneous) Gaussian ARMA(1,1) model:*

$$\xi_{n,t} = \mu + \gamma(\xi_{n,t-1} - \mu) + \eta\sqrt{1-\gamma^2}\varepsilon_t + \frac{1}{\sqrt{n}} I_{n,t}^{-1/2} u_t - \gamma \frac{1}{\sqrt{n}} I_{n,t-1}^{-1/2} u_{t-1}, \quad t = 1, \cdots, T, \tag{4.7}$$

where the observations are $\xi_{n,t} = \hat{f}_{n,t} + \frac{1}{2n} I_{n,t}^{-1} K_{n,t}^{(3)}$*, and the shocks* $(\varepsilon_t), (u_t)$ *are mutually independent* $IIN(0,1)$ *processes.*

The computation of the log-likelihood function of the ARMA(1,1) process (4.7) does not require the numerical inversion of a matrix of large dimension. Indeed, the (T,T) conditional variance-covariance matrix of $\xi_{n,t}$, $t = 1, \cdots, T$, is $\Omega_n = \eta^2(1-\gamma^2) Id_T + \frac{1}{n} B_n$, where B_n is the symmetric (T,T) matrix with elements equal to $I_{n,t}^{-1} + \gamma^2 I_{n,t-1}^{-1}$ in position (t,t), $-\gamma I_{n,t-1}^{-1}$ in positions $(t-1,t)$ and $(t,t-1)$, and zeros otherwise. At order $1/n$, we have:

$$\Omega_n^{-1} = \frac{1}{\eta^2(1-\gamma^2)} Id_T - \frac{1}{n\eta^4(1-\gamma^2)^2} B_n. \tag{4.8}$$

4.2 Approximate Linear Kalman Filter

We give below an equivalent statement of Proposition 5 in terms of an approximate linear Kalman filter.

PROPOSITION 6: *In a model with Gaussian autoregressive factor and macro-parameter* θ *only, a granularity adjusted maximum likelihood estimator of* θ *can be obtained by applying the standard Kalman filter to the linear Gaussian state space model with state equation:*

$$F_t = \mu + \gamma(F_{t-1} - \mu) + \eta\sqrt{1-\gamma^2}\varepsilon_t, \ \varepsilon_t \sim IIN(0,1), \tag{4.9}$$

and measurement equation:

$$\xi_{n,t} = F_t + \frac{1}{\sqrt{n}} I_{n,t}^{-1/2} u_t, \ u_t \sim IIN(0,1), \tag{4.10}$$

where $\xi_{n,t} = \hat{f}_{n,t} + \frac{1}{2n} I_{n,t}^{-1} K_{n,t}^{(3)}$*.*

By replacing F_t in (4.9) by its expression derived from (4.10), we recover the recursive equation (4.7) in Proposition 5. Equivalently, (4.9)-(4.10) is the linear state space representation of the ARMA(1,1) process of Proposition 5. The granularity adjustment in the measurement equation (4.10) concerns both the mean and the variance. Whereas the GA for variance corresponds to the usual asymptotic variance of $\hat{f}_{n,t}$, the GA for the mean is not correcting for the bias of $\hat{f}_{n,t}$ at order $1/n$. The reason is that the GA maximum likelihood estimator differs from the unfeasible maximum likelihood estimator of θ by a term of order smaller than $1/n$. The GA for mean is introduced to recover the bias at order $1/n$ of the unfeasible ML, which is not equal to zero. The estimator of macro-parameter θ in Proposition 6 computed with the linear Kalman filter differs numerically from the estimator in Proposition 5, when the latter is computed by using the approximate inverse variance-covariance matrix (4.8).

5 Granularity Adjustment for Value-at-Risk (VaR)

5.1 The Problem

The need for tractable approximation formulas in factor models with large cross-sectional size appeared first in Basel 2 regulation for credit risk [BCBS (2001)]. Let us consider a large homogenous portfolio of n financial risks. The total portfolio risk at $t+1$ can be written as:

$$W_{n,t+1} = \sum_{i=1}^{n} y_{i,t+1}, \tag{5.1}$$

where the individual risks $y_{i,t+1}$, $i = 1, \cdots, n$, are assumed to satisfy the assumptions of the nonlinear state space model in Section 2.1, with underlying factor F_{t+1}. For expository purpose, we include neither exogenous variables, nor lagged observations in the measurement equations. When the risk variables correspond to asset values, the VaR at risk level α, with $\alpha \in (0,1)$ and close to 0, is the opposite of the quantile of level α of the predictive distribution of $W_{n,t+1}$, called Profit and Loss (P&L) distribution. When the risk variables correspond to credit losses, the CreditVaR is computed for a confidence level $\alpha \in (0,1)$ close to 1, and corresponds to the α-quantile of the Loss and Profit (L&P) predictive distribution of $W_{n,t+1}$. In the sequel we focus on the second interpretation. It is usual to "standardize" the VaR by considering the VaR by individual asset, which corresponds to the (opposite of the) quantile at level α of $W_{n,t+1}/n$. This quantity $\text{VaR}_{n,t}(\alpha)$, say, depends on the portfolio size n and on the information $\mathbf{Y}_t$ available at time t. The VaR can be easily computed from the associated cumulative distribution function of $W_{n,t+1}/n$. Hence, we first focus on this function.

5.2 Approximation of the Predictive cdf of the Standardized Portfolio Risk

(i) By applying the Central Limit Theorem conditional on the factor value F_{t+1}, we can write for large n:

$$W_{n,t+1}/n = m(F_{t+1}) + \frac{\sigma(F_{t+1})}{\sqrt{n}} Z + O(1/n), \tag{5.2}$$

where:

$$m(f_{t+1}) = E[y_{i,t+1}|F_{t+1} = f_{t+1}], \quad \sigma^2(f_{t+1}) = V[y_{i,t+1}|F_{t+1} = f_{t+1}], \tag{5.3}$$

and Z is a standard Gaussian variable independent of $\mathbf{F}_{t+1}, \mathbf{Y}_t$. When n tends to infinity, the average of the individual risks tends to $m(F_{t+1})$, that is a stochastic variable. Indeed, due to systematic risk factor F_{t+1}, the risk cannot be entirely diversified by increasing the portfolio size. The terms $m(F_{t+1})$ and $\sigma(F_{t+1})$ (resp. Z) in the RHS of relation (5.2) show the effect of current systematic (resp. unsystematic) risks. Equation (5.2) provides an approximation of $W_{n,t+1}/n$ at order $o(1/n)$ for both the variance of $W_{n,t+1}/n$ and the bias. Indeed, the term $O(1/n)$ at order $1/n$ in expansion (5.2) is zero mean conditionally on $\mathbf{F}_{t+1}, \mathbf{Y}_t$, since $W_{n,t+1}/n$ is an unbiased estimator of $m(f_{t+1})$ conditional on $F_{t+1} = f_{t+1}$.

(ii) Let us now consider the cdf of $W_{n,t+1}/n$ given $\mathbf{F}_t, \mathbf{Y}_t, Z$. We have: [8]

$$\begin{aligned} P\left[W_{n,t+1}/n \leq w|\mathbf{F}_t, \mathbf{Y}_t, Z\right] &= \int \mathbb{1}_{m(f_{t+1}) + \frac{\sigma(f_{t+1})}{\sqrt{n}} Z \leq w} g(f_{t+1}|f_t) df_{t+1} + o(1/n) \\ &= a(w, f_t, Z/\sqrt{n}) + o(1/n), \text{ say.} \end{aligned} \tag{5.4}$$

Under mild regularity conditions, function $a(w, f, \varepsilon)$ is continuously differentiable w.r.t. the arguments f and ε at $\varepsilon = 0$ (see below). Function a summarizes the joint effect of lagged systematic risk and current unsystematic risk on the portfolio risk for large n.

(iii) We deduce that the predictive cdf $F_{n,t}(w) := P\left[W_{n,t+1}/n \leq w|\mathbf{Y}_t\right]$ of the standardized portfolio value given $\mathbf{Y}_t$ is:

$$\begin{aligned} F_{n,t}(w) &= E\left[a(w, F_t, Z/\sqrt{n})|\mathbf{Y}_t\right] + o(1/n) \\ &= E\left[a\left(w, \hat{f}_{n,t} + \frac{1}{n}\mu_{n,t} + \frac{1}{\sqrt{n}} I_{n,t}^{-1/2} Z^*, \frac{1}{\sqrt{n}} Z\right)|\mathbf{Y}_t\right] + o(1/n), \end{aligned} \tag{5.5}$$

[8] The fact that the remainder term in equation (5.4) is of order $o(1/n)$ is justified in Gagliardini, Gouriéroux (2010) by using that term $O(1/n)$ in expansion (5.2) is zero-mean.

where variable Z^* is standard Gaussian conditional on $\mathbf{Y}_t$, while $\mu_{n,t} = I_{n,t}^{-1} \frac{\partial \log g}{\partial f_t}(\hat{f}_{n,t}|\hat{f}_{n,t-1}) + \frac{1}{2} I_{n,t}^{-2} K_{n,t}^{(3)}$ is the GA mean for the filtering distribution and $I_{n,t}^{-1}/n$ the GA variance (see Proposition 1). Since the numerical Laplace approximation does not account for the stochastic feature of the observations, variables Z^* and Z are independent conditional on $\mathbf{Y}_t$ [9].

Then, we can expand equation (5.5) at order $1/n$. Since $E[Z] = E[Z^*]$ =0, $E[ZZ^*] = 0$, $E[Z^2] = E[(Z^*)^2] = 1$, we get:

$$\begin{aligned} F_{n,t}(w) &= a(w, \hat{f}_{n,t}, 0) + \frac{1}{n}\frac{\partial a}{\partial f}(w, \hat{f}_{n,t}, 0)\mu_{n,t} \\ &+ \frac{1}{2n}\left[I_{n,t}^{-1}\frac{\partial^2 a}{\partial f^2}(w, \hat{f}_{n,t}, 0) + \frac{\partial^2 a}{\partial \varepsilon^2}(w, \hat{f}_{n,t}, 0)\right] + o(1/n). \end{aligned} \tag{5.6}$$

In the above expression we distinguish three components:

The leading term:

$$a(w, \hat{f}_{n,t}, 0) = P\left[m(F_{t+1}) \leq w | F_t = \hat{f}_{n,t}\right] =: F_{\infty,t}(w), \tag{5.7}$$

is the cdf of $W_{n,t+1}/n$ evaluated at w and computed for a portfolio of infinite size, with perfect knowledge of the current factor value, identified with $\hat{f}_{n,t}$. Indeed, when $n = \infty$ the portfolio value per individual asset $W_{n,t+1}/n$ equals the individual conditional expected value $m(F_{t+1})$. Thus, the predictive cdf $F_{\infty,t}$ corresponds to the conditional distribution of $m(F_{t+1})$ given $F_t = \hat{f}_{n,t}$ [see e.g. Vasicek (1987, 1991) and Schoenbucher (2002) in a static framework, and Lamb, Perraudin, Van Landschoot (2008) in a dynamic framework]. This is the so-called Asymptotic Single Risk Factor (ASRF) model in Basel 2 terminology;

a first GA equal to $\frac{1}{2n}\frac{\partial^2 a}{\partial \varepsilon^2}(w, \hat{f}_{n,t}, 0)$ is introduced to account for the finite size of the portfolio, but still assuming a perfect knowledge of the current factor value;

the second GA, that is $\frac{1}{n}\frac{\partial a}{\partial f}(w, \hat{f}_{n,t}, 0)\mu_{n,t} + \frac{1}{2n} I_{n,t}^{-1}\frac{\partial^2 a}{\partial f^2}(w, \hat{f}_{n,t}, 0)$, takes into account the difference between the information sets $(\mathbf{F}_t, \mathbf{Y}_t)$ and $\mathbf{Y}_t$.

Due to the independence between Z and Z^*, there is no need for cross GA.

[9]More precisely, variable Z^* corresponds to the change of variable $Z^* = \sqrt{n} I_{n,t}^{1/2}(F_t - \hat{f}_{n,t} - \frac{1}{n}\mu_{n,t})$ in the conditional expectation w.r.t. F_t given $\mathbf{Y}_t$.

5.3 Granularity Adjustment of the standardized VaR

Finally, the GA of the VaR is directly deduced from (5.6) by applying the Bahadur's expansion. Let us denote by $Q_{n,t}$ (resp. $Q_{\infty,t}$) the quantile function corresponding to $F_{n,t}$ (resp. $F_{\infty,t}$), and assume that the density $f_{\infty,t}(w) = dF_{\infty,t}(w)/dw$ exists and is strictly positive. The quantile $Q_{\infty,t}$ and the pdf $f_{\infty,t}$ are called Cross-sectional Asymptotic (CSA) quantile and pdf, respectively. We have [Bahadur (1966)]:

$$Q_{n,t}(\alpha) - Q_{\infty,t}(\alpha) = -\frac{F_{n,t}\left[Q_{\infty,t}(\alpha)\right] - \alpha}{f_{\infty,t}\left[Q_{\infty,t}(\alpha)\right]} + o(1/n). \tag{5.8}$$

The GA for the quantile and for the standardized VaR are obtained by replacing $F_{n,t}$ and $F_{\infty,t}$ by their expressions using (5.6) and (5.7). In particular, the GA for the VaR is still at order $1/n$ and accounts for both the portfolio size and information effects discussed for the cdf.

Under suitable regularity conditions, the second-order partial derivative of function $a(w, f, \varepsilon)$ w.r.t. to ε at 0 can be expressed in terms of the conditional distributions defining the measurement and state equations. For instance, let us assume that function m is one-to-one. Then:

$$\begin{aligned}\frac{\partial^2 a}{\partial \varepsilon^2}(w, \hat{f}_{n,t}, 0) &= \frac{d}{dw}\left\{f_{\infty,t}(w) E\left[\sigma^2(F_{t+1}) | m(F_{t+1}) = w, F_t = \hat{f}_{n,t}\right]\right\} \\ &= \frac{d}{dw}\left\{f_{\infty,t}(w)\sigma^2[m^{-1}(w)]\right\}. \end{aligned} \tag{5.9}$$

This result is proved e.g. in Gagliardini, Gouriéroux (2010b), building on the local analysis of VaR in Gouriéroux, Laurent, Scaillet (2000) [see also Tasche (2000) and Martin, Wilde (2002)]. By combining equations (5.6), (5.8) and (5.9), we get the following Proposition:

PROPOSITION 7: *If function $m(.)$ is one-to-one, the VaR at risk level α is such that:*

$$VaR_{n,t}(\alpha) = Q_{\infty,t}(\alpha) + \frac{1}{n}[GA_{risk,t}(\alpha) + GA_{filt,t}(\alpha)],$$

where the GA for the finite portfolio size is:

$$GA_{risk,t}(\alpha) = -\frac{1}{2}\left\{\frac{d \log f_{\infty,t}(y)}{dy}\sigma^2[m^{-1}(y)] + \frac{d\sigma^2[m^{-1}(y)]}{dy}\right\}_{y=Q_{\infty,t}(\alpha)},$$

and the GA for filtering the current factor value is:

$$GA_{filt,t}(\alpha) = -\frac{1}{f_{\infty,t}\left[Q_{\infty,t}(\alpha)\right]}\left\{\mu_{n,t}\frac{\partial a}{\partial f}[Q_{\infty,t}(\alpha), \hat{f}_{n,t}, 0] + \frac{1}{2}I_{n,t}^{-1}\frac{\partial^2 a}{\partial f^2}[Q_{\infty,t}(\alpha), \hat{f}_{n,t}, 0]\right\}.$$

For a static factor model with $m(f) = f$, the GA for filtering the current factor value is equal to zero, and the GA for finite portfolio size becomes:

$$GA(\alpha) = -\frac{1}{2}\sigma^2[Q_\infty(\alpha)]\frac{d\log(f_\infty \cdot \sigma^2)}{dy}[Q_\infty(\alpha)],$$

where Q_∞ and f_∞ are the quantile and the pdf of F_t, respectively. This formula corresponds to the GA derived in Wilde (2001), Martin, Wilde (2002), Gordy (2003, 2004). Proposition 7 shows how the GA formula is extended and decomposed in models with a dynamic systematic factor.

5.4 Examples

Let us now derive the GA in two examples with exponential micro-density (see Section 3.2).

i) Linear Gaussian state space model

Let the variables $y_{i,t}$ follow the linear Gaussian state space model with measurement equations:

$$y_{i,t} = F_t + \sigma u_{i,t},\ i = 1, ..., n, \tag{5.10}$$

and state equation:

$$F_t = \mu + \gamma(F_{t-1} - \mu) + \eta\sqrt{1-\gamma^2}\varepsilon_t, \tag{5.11}$$

where $(u_{i,t})$, $i = 1, ..., n$, and (ε_t) are independent $IIN(0, 1)$ processes, and $|\gamma| < 1$. The factor F_t follows a stationary Gaussian AR(1) process, with a stationary distribution given by $N(\mu, \eta^2)$ and an autoregressive parameter equal to γ. The conditional distribution of $y_{i,t}$ given $F_t = f_t$ is Gaussian $N(f_t, \sigma^2)$, and hence the function $m(.)$ is given by $m(f) = f$, while the function $\sigma^2(f) = \sigma^2$ is constant. By using that the distribution of F_{t+1} conditional on $F_t = f_t$ is $N(\mu + \gamma(f_t - \mu), \eta^2(1 - \gamma^2))$, we deduce $a(w, f, 0) = \Phi\left(\frac{w - \mu - \gamma(f - \mu)}{\eta\sqrt{1-\gamma^2}}\right)$. By inversion w.r.t. w, we get the CSA VaR:

$$Q_{\infty,t}(\alpha) = \mu + \gamma(\hat{f}_{n,t} - \mu) + \eta\sqrt{1-\gamma^2}\Phi^{-1}(\alpha),$$

where the factor approximation $\hat{f}_{n,t} = \bar{y}_{n,t}$ is the cross-sectional average at date t. Let us now derive the GA's. From Proposition 7 the GA for the finite portfolio size is:

$$GA_{risk,t}(\alpha) = \frac{1}{2}\frac{\sigma^2}{\eta\sqrt{1-\gamma^2}}\Phi^{-1}(\alpha),$$

and the GA for the filtering of the factor value is given by:

$$GA_{filt,t}(\alpha) = \frac{\gamma\sigma^2}{\eta\sqrt{1-\gamma^2}}\left[\frac{1}{2}\gamma\Phi^{-1}(\alpha) - \hat{\varepsilon}_{n,t}\right],$$

where $\hat{\varepsilon}_{n,t} = \dfrac{\hat{f}_{n,t} - \mu - \gamma(\hat{f}_{n,t-1} - \mu)}{\eta\sqrt{1-\gamma^2}}$ denotes the standardized residual of the state equation.

ii) Nonlinear state space model for qualitative variables

Let us consider a portfolio of (zero-coupon) corporate bonds with maturity at $t+1$ and unitary nominal value, and denote by $y_{i,t+1}$ the issuer default indicators. Under the assumption of zero recovery rate, $W_{n,t+1}/n$ is the portfolio loss per individual loan at $t+1$. Let us assume that the dichotomous variables $y_{i,t+1}$ are such that $y_{i,t+1} = 1$, if $y^*_{i,t+1} < 0$, and $y_{i,t+1} = 0$, otherwise, where the latent variables $y^*_{i,t+1}$ correspond to the log of the asset-to-liability ratio of the issuers at date $t+1$. The variables $y^*_{i,t}$ are assumed to follow the linear Gaussian state space model (5.10)-(5.11). This defines a nonlinear state space model for dichotomous variables $y_{i,t}$. The measurement equation is such that the default indicator $y_{i,t}$ is Bernoulli distributed $\mathcal{B}(1, p_t)$ conditional on the factor value $F_t = f_t$, with conditional default probability:

$$p_t = P\left[y_{i,t} = 1 | F_t = f_t\right] = P\left[F_t + \sigma u_{i,t} < 0 | F_t = f_t\right] = \Phi\left(-f_t/\sigma\right). \tag{5.12}$$

Thus, the cross-sectional factor approximation at date t is given by $\hat{f}_{n,t} = -\sigma\Phi^{-1}(\bar{y}_{n,t})$, that is a nonlinear transformation of the cross-sectional default frequency $\bar{y}_{n,t}$, while functions $m(.)$ and $\sigma(.)$ are given by:

$$m(f_t) = \Phi\left(-f_t/\sigma\right), \quad \sigma^2(f_t) = \Phi\left(-f_t/\sigma\right)\left[1 - \Phi\left(-f_t/\sigma\right)\right].$$

Moreover, since the conditional probability of default p_t in (5.12) involves the ratio f_t/σ only, the distribution of the observable variables depends on three structural parameters, that are μ/σ, η/σ and γ.

Let us first compute the function $a(w, f, 0)$. Since function $m(.)$ is monotonically decreasing, we have:

$$\begin{aligned} a(w, f_t, 0) &= P[m(F_{t+1}) \leq w | F_t = f_t] = P\left[F_{t+1} \geq -\sigma\Phi^{-1}(w) | F_t = f_t\right] \\ &= \Phi\left(\frac{\sigma\Phi^{-1}(w) + \mu + \gamma(f_t - \mu)}{\eta\sqrt{1-\gamma^2}}\right). \end{aligned}$$

By inverting this function w.r.t. w and evaluating it at $f_t = \hat{f}_{n,t}$, we get the CSA VaR:

$$Q_{\infty,t}(\alpha) = \Phi\left(-\frac{\mu + \gamma(\hat{f}_{n,t} - \mu) + \eta\sqrt{1-\gamma^2}\Phi^{-1}(1-\alpha)}{\sigma}\right). \tag{5.13}$$

By the equivariance property of the VaR, the quantile $Q_{\infty,t}(\alpha)$ in (5.13) corresponds to the transformation by function $m(.)$ of the $(1-\alpha)$-quantile of the Gaussian distribution of F_{t+1} given $F_t = \hat{f}_{n,t}$. Let us now derive the GA of the quantile. From Proposition 7, the GA for finite portfolio size is:

$$GA_{risk,t}(\alpha) = \frac{1}{2}\left\{\frac{Q_{\infty,t}(\alpha)[1-Q_{\infty,t}(\alpha)]}{\phi\left(\Phi^{-1}[Q_{\infty,t}(\alpha)]\right)}\left(\frac{\sigma}{\eta\sqrt{1-\gamma^2}}\Phi^{-1}(\alpha) - \Phi^{-1}[Q_{\infty,t}(\alpha)]\right) + 2Q_{\infty,t}(\alpha) - 1\right\},$$

and the GA for filtering the current factor value is:

$$\begin{aligned} GA_{filt,t}(\alpha) &= \gamma\phi(\Phi^{-1}[Q_{\infty,t}(\alpha)])\frac{\bar{y}_{n,t}(1-\bar{y}_{n,t})}{(\phi[\Phi^{-1}(\bar{y}_{n,t})])^2}\left\{\frac{\sigma}{\eta\sqrt{1-\gamma^2}}\left(\hat{\varepsilon}_{n,t} - \frac{1}{2}\gamma\Phi^{-1}(1-\alpha)\right)\right. \\ &\qquad \left. + \frac{1-2\bar{y}_{n,t}}{\bar{y}_{n,t}\left(1-\bar{y}_{n,t}\right)}\phi[\Phi^{-1}(\bar{y}_{n,t})] - \frac{3}{2}\Phi^{-1}(\bar{y}_{n,t})\right\}. \end{aligned}$$

6 The Value of the Firm model with recovery

In this Section we consider a value of the firm model [Merton (1974), Vasicek (1991)] with single dynamic risk factor and non-zero recovery rate. We first introduce the model, then derive the cross-sectional approximation of the factor value and the filtering distribution, and finally compute the granularity adjustment of the portfolio VaR.

6.1 The model

Let $A_{i,t}$ and $L_{i,t}$ denote the asset of firm i at date t, and the firm liability maturing at date t, respectively. The percentage loss of the debt holder at date t is:

$$y_{i,t} = \mathbb{1}_{A_{i,t}<L_{i,t}}\left(1 - \frac{A_{i,t}}{L_{i,t}}\right) = \left(1 - \frac{A_{i,t}}{L_{i,t}}\right)^{+}. \tag{6.1}$$

The loss variable $y_{i,t}$ is the product of the default indicator $\mathbb{1}_{A_{i,t}<L_{i,t}}$, 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 $1 - \frac{A_{i,t}}{L_{i,t}}$. [10] At a given date t and for given liability $L_{i,t}$, the second equality in (6.1) corresponds to the interpretation of the loss $L_{i,t}y_{i,t}$ incurred by the debt holder as the payoff of a put option written on the firm asset with strike equal to the liability [Merton (1974)].

Let us assume that the log asset/liability ratios of the firms follow a linear single risk factor (SRF) model:

$$\log\left(\frac{A_{i,t}}{L_{i,t}}\right) = F_t + \sigma u_{i,t}, \tag{6.2}$$

where F_t is a systematic risk factor common across firms, $u_{i,t} \sim IIN(0,1)$ are unsystematic (idiosyncratic) risks independent over time and across firms, and independent of factor (F_t), and σ is the unsystematic (idiosyncratic) volatility. Factor (F_t) follows a stationary Gaussian AR(1) process:

$$F_t = \mu + \gamma(F_{t-1} - \mu) + \eta\sqrt{1-\gamma^2}\varepsilon_t, \tag{6.3}$$

where the innovations are $\varepsilon_t \sim IIN(0,1)$ and the autoregressive coefficient γ is such that $|\gamma| < 1$. Parameters μ and η are the mean and volatility of the stationary distribution of F_t, respectively. The specification (6.2)-(6.3) extends the SRF model introduced by Vasicek (1991) and considered in Basel 2 regulation [BCBS (2001)] to a dynamic framework. The unobservable factor F_t has a linear effect on the latent asset/liability ratios by means of the drift only. However, our interest is in the individual risks $y_{i,t}$. Conditional on factor F_t, the mean and variance of $y_{i,t}$ depend on the factor. Thus, we get a model for the observable variables with both stochastic mean $m(F_t)$ and stochastic volatility $\sigma(F_t)$ (see Section 6.3).

The dynamic SRF model involves 4 structural parameters. As usual, it is interesting to introduce an alternative parameterization, which is easier for interpretation and calibration purposes.

[10] The results in this section are easily extended to the model $y_{i,t} = \mathbf{1}_{A_{i,t}<L_{i,t}}\left(1 - \delta\frac{A_{i,t}}{L_{i,t}}\right)$, where δ is a parameter such that $0 \leq \delta \leq 1$ [Eom, Helwege, Huang (2004)]. In this model, when the firm is in default and the assets are liquidated, only the part $\delta A_{i,t}$ can be recovered by the debt holder, and the liquidation cost $(1-\delta)A_{i,t}$ is lost. When $\delta = 1$ we get model (6.1), while in the other extreme case $\delta = 0$, we get the standard Value of the Firm model with pure default and zero recovery rate [see Example ii) in Section 6.4].

The unconditional probability of default PD and asset correlation ρ are given by: [11]

$$PD = P\left[\log\left(A_{i,t}/L_{i,t}\right) < 0\right] = \Phi\left(-\frac{\mu}{\sqrt{\eta^2+\sigma^2}}\right), \tag{6.4}$$

and:

$$\rho = corr\left[\log\left(A_{i,t}/L_{i,t}\right), \log\left(A_{j,t}/L_{j,t}\right)\right] = \frac{\eta^2}{\eta^2+\sigma^2}, \tag{6.5}$$

for $i \neq j$, respectively. Moreover, the unconditional expected (percentage) loss given default (ELGD) is defined by:

$$ELGD = E\left[1 - \frac{A_{i,t}}{L_{i,t}} | \frac{A_{i,t}}{L_{i,t}} < 1\right]. \tag{6.6}$$

Since $ELGD \cdot PD = E\left[(1 - A_{i,t}/L_{i,t})^+\right]$ and $\log(A_{i,t}/L_{i,t}) \sim N(\mu, \eta^2+\sigma^2)$, we deduce that $ELGD \cdot PD$ is equal to the price of a put option in the Black-Scholes model with volatility parameter $\sqrt{\eta^2+\sigma^2}$ and risk-free rate $\mu + \frac{1}{2}(\eta^2+\sigma^2)$, divided by the price of the zero-coupon bond at the same maturity, and we get [see Geske (1977) and Appendix 6]:

$$ELGD = 1 - \exp\left[\mu + \frac{1}{2}(\eta^2+\sigma^2)\right] \frac{\Phi\left(-\frac{\mu}{\sqrt{\eta^2+\sigma^2}} - \sqrt{\eta^2+\sigma^2}\right)}{\Phi\left(-\frac{\mu}{\sqrt{\eta^2+\sigma^2}}\right)}. \tag{6.7}$$

Equations (6.4), (6.5) and (6.7) define a one-to-one mapping between structural parameters (μ, η, σ) and parameters $(PD, \rho, ELGD)$. Indeed, we have (see Appendix 6):

$$\mu = -\tau\Phi^{-1}(PD), \quad \eta = \tau\sqrt{\rho}, \quad \sigma = \tau\sqrt{1-\rho}, \tag{6.8}$$

where $\tau \geq 0$ is the unique solution of the equation:

$$PD - \exp\left[\frac{1}{2}\tau^2 - \Phi^{-1}(PD)\tau\right]\Phi\left[\Phi^{-1}(PD) - \tau\right] = ELGD \cdot PD. \tag{6.9}$$

The LHS of equation (6.9) is the Black-Scholes put option price as a function of volatility $\tau = \sqrt{\eta^2+\sigma^2}$ and for given risk-neutral probability PD that the put is in the money at maturity. Thus, the solution τ of equation (6.9) is similar to an implied volatility. Note that both τ and μ depend

[11] These summary statistics have to be distinguished from their conditional counterparts given the observed histories of individual risks. The latter ones are path dependent due to the unobservability of the factor and the nonlinear dependence of the individual risks $(y_{i,t})$ with respect to the factor. Thus, model (6.1)-(6.3) implies both conditional heteroscedasticity and dynamic conditional correlation in the underlying log asset/liability ratios.

on PD and $ELGD$ only. To summarize, the dynamic SRF model can be parameterized in terms of unconditional probability of default PD, asset correlation ρ, expected loss given default $ELGD$, and the autoregressive coefficient of the factor γ.

The one-to-one relationship between structural parameters (μ, η, σ) and reduced form parameters $(PD, \rho, ELGD)$ is especially important for calibration. Indeed, historical estimates of PD, ρ and $ELGD$ are easily obtained in practice and, by inverting the relationship, we deduce estimates of the structural parameters. As an illustration, we give in Table 2 the values of the structural parameters corresponding to some values of the reduced form parameters suggested by the Basel Committee [see BCBS (2001)], i.e. Basel implied structural parameters. Specifically, the values 0.45 and 0.75 for $ELGD$ correspond to senior classes on corporate, sovereigns and banks not secured, and subordinated classes on corporate, sovereigns and banks, respectively. The values 0.12 and 0.24 for ρ are the minimal and maximal asset correlations considered in Basel 2, respectively, for debt without guarantees, while $\rho = 0.50$ is the value of asset correlation for guaranteed debt. The values 1.5% and 5% for PD are representative for yearly default probabilities of obligors in rating classes BB and B in Fitch, respectively. Some of the parameter values in Table 2 are used in the illustrations of the next subsections.

6.2 Cross-sectional factor approximation and filtering distribution

Let us first write the dynamic SRF model as a nonlinear state space model. From equations (6.1) and (6.2) the loss variable is such that:

$$y_{i,t} = [1 - \exp(F_t + \sigma u_{i,t})]^+ .$$

Thus, the measurement equations correspond to a Gaussian tobit regression model with endogenous variable $\log(1 - y_{i,t})$, mean f_t and variance σ^2, and are characterized by the conditional density [Tobin (1958)]:

$$\prod_{i=1}^{n} h(y_{i,t}|f_t) = \prod_{i:y_{i,t}>0} \left[\frac{1}{\sigma}\phi\left(\frac{\log(1-y_{i,t}) - f_t}{\sigma}\right)\frac{1}{1-y_{i,t}}\right] \prod_{i:y_{i,t}=0} \Phi(f_t/\sigma), \qquad (6.10)$$

while the state equation is the Gaussian AR(1) model (6.3).

Let us now compute the cross-sectional factor approximation and derive the approximate filtering distribution of the unobservable factor value. The cross-sectional maximum likelihood approximation of the factor value at date t is given by:

$$\hat{f}_{n,t} = \arg\max_{f_t} \left\{ -\frac{1}{2\sigma^2} \sum_{i:y_{i,t}>0} [\log(1-y_{i,t}) - f_t]^2 + (n-n_t)\log\Phi(f_t/\sigma) \right\}, \tag{6.11}$$

where $n_t = \sum_{i=1}^{n} \mathbb{1}_{y_{i,t}>0}$ denotes the number of defaults at date t. The factor approximation $\hat{f}_{n,t}$ is the solution of the nonlinear first-order condition:

$$\frac{1}{\sigma} \sum_{i:y_{i,t}>0} [\log(1-y_{i,t})] - n_t(\hat{f}_{n,t}/\sigma) + (n-n_t)\lambda(\hat{f}_{n,t}/\sigma) = 0,$$

where:

$$\lambda(x) = \frac{\phi(x)}{\Phi(x)}, \tag{6.12}$$

denotes the inverse Mill's ratio. From Proposition 1 the approximate filtering distribution of F_t is Gaussian with density (see Appendix 6):

$$N\left(\hat{f}_{n,t} + \frac{1}{n}\left[-I_{n,t}^{-1} \frac{\hat{f}_{n,t} - \mu - \gamma(\hat{f}_{n,t-1} - \mu)}{\eta^2(1-\gamma^2)} + \frac{1}{2} I_{n,t}^{-2} K_{n,t}^{(3)} \right], \frac{1}{n} I_{n,t}^{-1} \right), \tag{6.13}$$

where the quantities $I_{n,t} = -K_{n,t}^{(2)}$ and $K_{n,t}^{(3)}$ are given by:

$$I_{n,t} = \frac{1}{\sigma^2}\left\{ \frac{n_t}{n} + \left(1 - \frac{n_t}{n}\right) \lambda\left(\hat{f}_{n,t}/\sigma\right) \left[\hat{f}_{n,t}/\sigma + \lambda\left(\hat{f}_{n,t}/\sigma\right)\right] \right\}, \tag{6.14}$$

and:

$$K_{n,t}^{(3)} = -\frac{1}{\sigma^3}\left(1 - \frac{n_t}{n}\right) \lambda\left(\hat{f}_{n,t}/\sigma\right) \left\{ 1 - \left[\hat{f}_{n,t}/\sigma + \lambda\left(\hat{f}_{n,t}/\sigma\right)\right] \left[\hat{f}_{n,t}/\sigma + 2\lambda\left(\hat{f}_{n,t}/\sigma\right)\right] \right\}, \tag{6.15}$$

respectively. The quantities $I_{n,t}$ and $K_{n,t}^{(3)}$ depend on the information at date t through the factor approximation $\hat{f}_{n,t}$ and the default frequency n_t/n only. Moreover, the different quantities $\hat{f}_{n,t}$, $I_{n,t}$ and $K_{n,t}^{(3)}$ involve parameter σ only. The other structural parameters μ, η and γ impact the filtering distribution through the standardized residual $\hat{\varepsilon}_{n,t} = \frac{\hat{f}_{n,t} - \mu - \gamma(\hat{f}_{n,t-1} - \mu)}{\eta\sqrt{1-\gamma^2}}$ and the conditional standard deviation $\eta\sqrt{1-\gamma^2}$. [12]

[12] We have noted in Section 6.1 that parameters σ, μ and η are easily calibrated (see e.g. Table 2). Thus, the factor value at date t can be estimated by considering the cross-sectional ML estimator of f_t given in (6.11) with σ replaced by its calibrated approximation, $\hat{f}_{n,t}^*$, say. Then, the remaining structural parameter γ is easily deduced from the historical correlation between $\hat{f}_{n,t}^*$ and $\hat{f}_{n,t-1}^*$.

In Figure 1 we display the conditional distribution of F_t given $F_{t-1} = \mu$ and the approximate filtering distribution of F_t for different values of the cross-sectional dimension n, that are $n = 50$, 100 and 1000. The micro-information is such that the cross-sectional factor approximations are $\hat{f}_{n,t} = \hat{f}_{n,t-1} = \mu$ and the default frequency is $n_t/n = PD$. The structural parameters are such that $ELGD = 0.45$, $PD = 5\%$, $\rho = 0.12$ and $\gamma = 0.5$ (see Table 2). When n gets larger, the filtering distribution features a smaller variance and peaks at the cross-sectional factor approximation $\hat{f}_{n,t}$, as an effect of the increasing micro-information. The mean of the approximate filtering distributions differs from $\hat{f}_{n,t}$ because of the bias adjustment.

In Figure 2 we investigate the effect of the micro-information on the mean and the standard deviation of the filtering distribution of F_t for $n = 100$. When $n_t/n = PD$ and $\hat{\varepsilon}_{n,t} = 0$, the mean of the filtering distributions is close to $\hat{f}_{n,t}$ and the standard deviation is increasing w.r.t. $\hat{f}_{n,t}$ (upper left Panel). When $\hat{f}_{n,t} = \mu$ and $\hat{\varepsilon}_{n,t} = 0$, the filtering distribution of the factor F_t is not very sensitive to the default frequency n_t/n (upper right Panel). Finally, when $\hat{f}_{n,t} = \mu$ and $n_t/n = PD$, the mean of the filtering distribution is decreasing w.r.t. the standardized residual $\hat{\varepsilon}_{n,t}$ (lower left Panel). For given cross-sectional factor approximation $\hat{f}_{n,t}$, the mean of the filtering distribution is larger (resp., smaller) than $\hat{f}_{n,t}$ when $\hat{\varepsilon}_{n,t} < c$ (resp., $\hat{\varepsilon}_{n,t} > c$), where $c = \frac{1}{2}\eta\sqrt{1-\gamma^2} I_{n,t}^{-1} K_{n,t}^{(3)}$ is close to zero. Since the coefficient of $\hat{f}_{n,t}$ in the mean of the filtering distribution of F_t is $1 - \frac{1}{n}\frac{I_{n,t}^{-1}}{\eta^2(1-\gamma^2)} < 1$, a cross-sectional shock in $\hat{f}_{n,t}$ at date t is transmitted less than fully to the mean of the filtering distribution of F_t. This effect is more pronounced when the autoregressive coefficient is large and close to 1 (lower right Panel).

6.3 The granularity adjustment for portfolio VaR

Let us first derive the functions $m(f_{t+1})$ and $\sigma^2(f_{t+1})$ [see equations (5.3) in Section 5.2]. Conditional on the future value of the factor F_{t+1}, the loss variable:

$$y_{i,t+1} = [1 - \exp(F_{t+1} + \sigma u_{i,t+1})]^+,$$

corresponds to the payoff of a European put option with strike 1, time-to-maturity 1 and current value of the underlying asset equal to $\exp(F_{t+1})$, in the Black-Scholes model with volatility parameter σ and risk-free rate $\sigma^2/2$. Then, we have (see Lemma 1 in Appendix 6):

$$\begin{aligned} m(f_{t+1}) &= E[\mathbb{1}_{u_{i,t+1}<-F_{t+1}/\sigma}(1-\exp(F_{t+1}+\sigma u_{i,t+1}))|F_{t+1}=f_{t+1}] \\ &= \Phi(-f_{t+1}/\sigma)-\exp\left(f_{t+1}+\frac{\sigma^2}{2}\right)\Phi\left(-f_{t+1}/\sigma-\sigma\right), \end{aligned} \tag{6.16}$$

which corresponds to the Black-Scholes price of the put option divided by the zero-coupon bond price of the same maturity. The function m is monotone decreasing, since the variable $y_{i,t+1}$ is decreasing w.r.t. F_{t+1}. To compute the derivative of function m, write $m(f_{t+1}) = \int_{-\infty}^{-f_{t+1}/\sigma}[1 - \exp(f_{t+1}+\sigma u)]\phi(u)du$. Then, the derivative of function m is given by:

$$\frac{dm(f_{t+1})}{df_{t+1}} = -\exp(f_{t+1})\int_{-\infty}^{-f_{t+1}/\sigma}\exp(\sigma u)\phi(u)du = -\exp\left(f_{t+1}+\frac{\sigma^2}{2}\right)\Phi\left(-f_{t+1}/\sigma-\sigma\right)\leq 0, \tag{6.17}$$

which is the delta of the put multiplied by $\exp(f_{t+1})$. The function $\sigma^2(f_{t+1})$ is given by (see Appendix 6):

$$\begin{aligned} \sigma^2(f_{t+1}) &= m(f_{t+1})[1-m(f_{t+1})]-\exp\left(f_{t+1}+\frac{\sigma^2}{2}\right)\Phi\left(-f_{t+1}/\sigma-\sigma\right) \\ &\quad +\exp(2f_{t+1}+2\sigma^2)\Phi\left(-f_{t+1}/\sigma-2\sigma\right). \end{aligned} \tag{6.18}$$

This is the variance of the payoff of a short-term put option with strike 1 and underlying asset price $\exp(F_{t+1})$ in the Black-Scholes model. Finally, the derivative of $\sigma^2(f_{t+1})$ w.r.t. f_{t+1} is given by (see Appendix 6):

$$\begin{aligned} \frac{d\sigma^2(f_{t+1})}{df_{t+1}} &= -2\exp\left(f_{t+1}+\frac{\sigma^2}{2}\right)\Phi\left(-f_{t+1}/\sigma-\sigma\right)[1-\Phi(-f_{t+1}/\sigma)] \\ &\quad +2\exp\left(2f_{t+1}+2\sigma^2\right)\Phi\left(-f_{t+1}/\sigma-2\sigma\right)-2\exp\left(2f_{t+1}+\sigma^2\right)\left[\Phi\left(-f_{t+1}/\sigma-\sigma\right)\right]^2. \end{aligned} \tag{6.19}$$

Functions $m(f_{t+1})$ and $\sigma^2(f_{t+1})$ as well as their first-order derivatives involve micro-parameter σ. These functions are displayed in Figure 3 for a value of σ corresponding to $ELGD = 0.45$, $PD = 5\%$ and $\rho = 0.12$ (see Table 2).

Let us now compute the function $a(w, \hat{f}_{n,t}, 0)$ [see equation (5.7) in Section 5.2]. We have:

$$a(w,\hat{f}_{n,t},0) = P[m(F_{t+1})\leq w|F_t=\hat{f}_{n,t}] = P[F_{t+1}\geq m^{-1}(w)|F_t=\hat{f}_{n,t}] \tag{6.20}$$

$$= \Phi\left(-\frac{m^{-1}(w)-\mu-\gamma(\hat{f}_{n,t}-\mu)}{\eta\sqrt{1-\gamma^2}}\right), \tag{6.21}$$

where m^{-1} denotes the inverse of function m and we used that m is monotone decreasing. The conditional pdf of $m(F_{t+1})$ given $F_t = \hat{f}_{n,t}$ is obtained by differentiating function a w.r.t. w and is given by:

$$\begin{aligned} f_{\infty,t}(w) &= \frac{\partial a(w, \hat{f}_{n,t}, 0)}{\partial w} = -\frac{1}{\eta\sqrt{1-\gamma^2}}\phi\left(-\frac{m^{-1}(w) - \mu - \gamma(\hat{f}_{n,t} - \mu)}{\eta\sqrt{1-\gamma^2}}\right)\frac{dm^{-1}(w)}{dw} \\ &= \frac{1}{\eta\sqrt{1-\gamma^2}}\phi\left(\frac{m^{-1}(w) - \mu - \gamma(\hat{f}_{n,t} - \mu)}{\eta\sqrt{1-\gamma^2}}\right)\frac{\exp\left(-m^{-1}(w) - \frac{\sigma^2}{2}\right)}{\Phi\left(-\frac{m^{-1}(w)}{\sigma} - \sigma\right)}, \end{aligned} \quad (6.22)$$

where we used (6.17).

The next Proposition 8 gives the CSA VaR and the GA in the Value of the Firm model with dynamic factor and non-zero recovery rate (see Appendix 6 for the proof).

PROPOSITION 8: (i) *The CSA VaR at confidence level α is given by:*

$$Q_{\infty,t}(\alpha) = m\left[Q^*_{\infty,t}(1-\alpha)\right], \quad (6.23)$$

where:

$$Q^*_{\infty,t}(1-\alpha) = \mu + \gamma(\hat{f}_{n,t} - \mu) + \eta\sqrt{1-\gamma^2}\Phi^{-1}(1-\alpha), \quad (6.24)$$

function m is defined in (6.16), and $\hat{f}_{n,t}$ is the cross-sectional factor approximation in (6.11).

(ii) *The granularity adjustment is $\frac{1}{n}[GA_{risk,t}(\alpha) + GA_{filter,t}(\alpha)]$, where the GA for risk is:*

$$\begin{aligned} GA_{risk,t}(\alpha) &= -\frac{1}{2}\frac{1}{\frac{dm}{df_{t+1}}[Q^*_{\infty,t}(1-\alpha)]}\left\{\left(\frac{1}{\eta\sqrt{1-\gamma^2}}\Phi^{-1}(\alpha) + \frac{1}{\sigma}\lambda\left[-\frac{Q^*_{\infty,t}(1-\alpha)}{\sigma} - \sigma\right] - 1\right)\right. \\ &\left.\cdot\sigma^2[Q^*_{\infty,t}(1-\alpha)] + \frac{d\sigma^2}{df_{t+1}}[Q^*_{\infty,t}(1-\alpha)]\right\}, \end{aligned} \quad (6.25)$$

and the GA for filtering is:

$$GA_{filt,t}(\alpha) = -\gamma\frac{dm}{df_{t+1}}[Q^*_{\infty,t}(1-\alpha)]I^{-1}_{n,t}\left\{\frac{1}{\eta\sqrt{1-\gamma^2}}\left(\hat{\varepsilon}_{n,t} - \frac{1}{2}\gamma\Phi^{-1}(1-\alpha)\right) - \frac{1}{2}I^{-1}_{n,t}K^{(3)}_{n,t}\right\}, \quad (6.26)$$

and where functions $\lambda[.]$, $dm[.]/df_{t+1}$, $\sigma^2[.]$ and $d\sigma^2[.]/df_{t+1}$ are given in (6.12), (6.17), (6.18) and (6.19), respectively, the summary statistics $I_{n,t}$ and $K^{(3)}_{n,t}$ are given in (6.14) and (6.15), and $\hat{\varepsilon}_{n,t} = \frac{\hat{f}_{n,t} - \mu - \gamma(\hat{f}_{n,t-1} - \mu)}{\eta\sqrt{1-\gamma^2}}$.

By the equivariance property of the quantile function under monotone decreasing transformations, the α-quantile $VaR_{\infty,t}(\alpha)$ is the transformation by function m of the $(1-\alpha)$-quantile $Q^*_{\infty,t}(1-\alpha)$ of the Gaussian distribution of F_{t+1} given $F_t = \hat{f}_{n,t}$. The CSA VaR depends on the unconditional mean μ and volatility η of the systematic factor, on its autoregressive coefficient γ, as well as on the factor approximation $\hat{f}_{n,t}$, through the Gaussian quantile $Q^*_{\infty,t}(1-\alpha)$. It depends on the idiosyncratic volatility parameter σ through transformation m. The GA for risk involves parameters μ, η, σ and γ and depends on the information through the Gaussian quantile $Q^*_{\infty,t}(1-\alpha)$ only. Similarly, the GA for filtering involves the four structural parameters and depends on the information through $Q^*_{\infty,t}(1-\alpha)$, the standardized residual $\hat{\varepsilon}_{n,t}$ and quantities $I_{n,t}$ and $K^{(3)}_{n,t}$.

In Figure 4 we display the CSA VaR, the risk and filtering components of the GA, as well as the GA VaR for $n = 100$ and $n = 1000$, as functions of the cross-sectional factor approximation $\hat{f}_{n,t}$. The default frequency at date t is $n_t/n = PD$ and the lagged value of the cross-sectional factor approximation is $\hat{f}_{n,t-1} = \mu$. The parameters are such that $ELGD = 0.45$, $PD = 5\%$, $\rho = 0.12$ and $\gamma = 0.5$ (see Table 2). The confidence level is $\alpha = 0.995$. The CSA VaR is decreasing w.r.t. $\hat{f}_{n,t}$, since larger factor values imply larger asset/liability ratios. The patterns of the GA components for risk and filtering are very different. The GA for risk admits positive values and is decreasing w.r.t. $\hat{f}_{n,t}$ over the displayed range of factor values $[\mu - 3\eta, \mu + 3\eta]$, while the GA for filtering admits both positive and negative values and is increasing w.r.t. $\hat{f}_{n,t}$. Indeed, when $\hat{f}_{n,t}$ is large, the standardized residual $\hat{\varepsilon}_{n,t}$ is also large and positive, and thus the mean of the approximate filtering distribution is smaller than $\hat{f}_{n,t}$ (see Figure 2). This granularity adjustment in the filtering distribution implies a less optimist factor value at date t compared to $\hat{f}_{n,t}$, which yields an upward adjustment for the portfolio VaR. For a portfolio of $n = 100$ contracts, the GA is large and relevant for most values of the cross-sectional factor approximation $\hat{f}_{n,t}$. For $n = 1000$, the GA is about 5%-10% of the CSA VaR for moderate to large values of $\hat{f}_{n,t}$, and is mostly due to the filtering of the unobservable factor value.

In Figure 5 we display the CSA VaR, the GA VaR and the risk and filtering GA components for $PD = 1.5\%$. Compared to Figure 4, the CSA VaR is smaller for the corresponding factor values, the GA for risk is slightly smaller and the GA for filtering is larger. This results in granularity adjustments that are very large for portfolio size $n = 100$, and about 20% of the CSA VaR for moderate factor values when $n = 1000$.

The last Figures 6 and 7 illustrate the dynamic features of CSA and GA VaR. The third Panel in Figure 6 provides a simulated path of the factor and its cross-sectional approximation. We observe that $\hat{f}_{n,t}$ has some tendency to smooth the underlying factor values. The two upper panels are describing the evolution of the losses with zero and non-zero recovery rates. When the non-zero recovery rate is taken into account, the loss is smaller and smoother. The corresponding evolution of the VaR measures and their components are displayed in Figure 7. The GA VaR is larger and smoother 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. Table 3 displays the (cross-) autocorrelograms of the CSA and GA VaR series computed by Monte-Carlo simulation. The GA VaR series is more persistent.

Finally, it is necessary to check if the GA VaR is preferable to the CSA VaR in terms of the frequency and dynamic pattern of violations. In Table 4 we report the values of different summary statistics associated with predictability test procedures [Giacomini, White (2006)]. More precisely, the conditional VaR satisfies the conditional moment restriction:

$$P\left[W_{n,t}/n \leq VaR_{n,t}(\alpha)|\mathbf{Y}_t\right] = \alpha \quad \Leftrightarrow \quad E\left[\mathbb{1}_{W_{n,t}/n \geq VaR_{n,t}(\alpha)} - (1-\alpha)|\mathbf{Y}_t\right] = 0.$$

Thus, a battery of specification tests can be introduced by considering the unconditional moment restrictions:

$$E\left[\xi_t\left(\mathbb{1}_{W_{n,t}/n \geq VaR_{n,t}(\alpha)} - (1-\alpha)\right)\right] = 0,$$

where ξ_t is a selected instrument function of the information $\mathbf{Y}_t$. We provide in Table 4 the values of different such statistics, computed by Monte-Carlo. When the instrument is constant $\xi_t = 1$ (second row of Table 4), we get the standard criterion for ex-post validation in Basel 2, that corresponds to the frequency of violations in excess of the nominal risk level $1-\alpha$. Other instrumental variables are selected in rows 3-8 of Table 4, and the results are displayed in terms of correlation between $\mathbb{1}_{W_{n,t}/n \geq VaR_{n,t}(\alpha)} - (1-\alpha)$ and these instruments. It is immediately seen that the values of the summary statistics are significantly smaller in absolute value for the GA VaR.

7 Concluding Remarks

Recently there have been several developments in the literature on nonlinear factor models with individual observations and macro-factors. These developments are especially relevant in Finance and Insurance when large homogenous portfolios of individual contracts, such as loans, mortgages, revolving credits, Credit Default Swaps, life insurance contracts, are involved. This paper shows how the difficulties encountered with nonlinear Kalman recursions can be solved by an appropriate use of the micro-information. The granularity principle followed in this paper consists in expanding the quantity of interest with respect to $1/n$, where n is the cross-sectional dimension. The term of order 0 in $1/n$ corresponds to the Asymptotic Single Risk Factor model, that is, to the virtual case of an infinite cross-sectional size; the next term of order $1/n$ provides the granularity adjustment. We have seen that this principle works for rather different quantities of interest such as a filtering distribution, a predictive distribution, the maximum likelihood estimator of a macro-parameter, or the VaR of a large homogenous portfolio.

References

[1] Bahadur, R. (1966): “A Note on Quantiles in Large Samples”, Annal of Mathematical Statistics“, 37, 577-580.

[2] Basel Committee on Banking Supervision (2001): ”The New Basel Capital Accord”, Consultative Document of the Bank for International Settlements.

[3] Bates, D. (2009): “U.S. Stock Market Crash Risk, 1926-2006”, Working Paper.

[4] Cappé, O., Moulines, E., and T., Rydén (2005): *Inference in Hidden Markov Models*, Springer.

[5] Cipra, T., and A., Rubio (1991): ”Kalman Filter with a Nonlinear Non-Gaussian Observation Relation”, Trabajos de Estadistica, 6, 111-119.

[6] Eom, Y., Helwege, J., and J., Huang (2004): “Structural Models of Corporate Bond Pricing: An Empirical Analysis”, Review of Financial Studies, 17, 499-544.

[7] Gagliardini, P., and C., Gouriéroux (2009): ”Approximate Derivative Pricing in Large Classes of Homogeneous Assets with Systematic Risk”, CREST DP.

[8] Gagliardini, P., and C., Gouriéroux (2010a): “Efficiency in Large Dynamic Panel Models with Common Factor”, CREST DP.

[9] Gagliardini, P., and C., Gouriéroux (2010b): “ Granularity Adjustment for Risk Measures: Systematic vs Unsystematic Risks”, CREST DP.

[10] Geske, R. (1977): “The Valuation of Corporate Liabilities as Compound Options”, Journal of Financial and Quantitative Analysis, 12, 541-552.

[11] Giacomini, R., and H., White (2006): “Tests of Conditional Predictive Ability”, Econometrica, 74, 1545-1578.

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

[13] Gordy, M. (2004): “Granularity Adjustment in Portfolio Credit Risk Measurement”, in G. Szego, ed., Risk Measures for the 21.st Century, Wiley.

[14] Gouriéroux, C., and J., Jasiak (2002): ”State Space Models with Finite Dimensional Dependence”, Journal of Time Series Analysis, 22, 665-678.

[15] Gouriéroux, C., Laurent, J.-P., and O., Scaillet (2000): “Sensitivity Analysis of Values at Risk”, Journal of Empirical Finance, 7, 225-245.

[16] Gouriéroux, C., and A., Monfort (2009): ”Granularity in Qualitative Factor Model”, Journal of Credit Risk, 5, 29-61.

[17] Hamilton, J. (1989): “A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle”, Econometrica, 57, 357-384.

[18] Jensen, J. (1995): *Saddlepoint Approximations*, Clarendon Press, Oxford.

[19] Kalman, R. (1960): ”A New Approach to Linear Filtering and Prediction Problems”, Journal of Basic Engineering, 82, 34-35.

[20] Kalman, R., and R., Bucy (1961): ”New Results in Linear Filtering and Prediction Theory”, Journal of Basic Engineering, 83, 95-108.

[21] Kitagawa, G. (1987): ”Non-Gaussian State Space Modelling of Nonstationary Time Series”, Journal of the American Statistical Association, 82, 1032-1063.

[22] Kitagawa, G. (1996): ”Monte-Carlo Filter and Smoother for Non-Gaussian Nonlinear State Space Models”, Journal of Computational and Graphical Statistics, 5, 1-25.

[23] Lamb, R., Perraudin, W., and A., Van Landschoot (2008): ”Dynamic Pricing of Synthetic Collateralized Debt Obligations”, Imperial College Working Paper.

[24] Martin, R., and T., Wilde (2002): ”Unsystematic Credit Risk”, Risk, 15, 123-128.

[25] Masreliez, C. (1975): ”Approximate Non-Gaussian Filtering with Linear State and Observation Relations”, IEEE Transactions on Automatic Control, 20, 107-110.

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

[27] Pitt, M., and N., Shephard (2001): “Filtering via Simulation: Auxiliary Particle Filters”, Journal of the American Statistical Association, 94, 590-599.

[28] Schick, I., and S., Mitter (1994): “Robust Recursive Estimation in the Presence of Heavy-tailed Observation Noise”, Annals of Statistics, 22, 1045-1080.

[29] Schoenbucher, P. (2002): ”Taken to the Limit: Simple and Not-so-simple Loan Loss Distributions”, University of Bonn Working Paper.

[30] Tasche, D. (2000): ”Conditional Expectation as Quantile Derivative”, Bundesbank, DP.

[31] Tobin, J. (1958): ”Estimation of Relationships for Limited Dependent Variables”, Econometrica, 26, 24-36.

[32] Vasicek, O. (1987): ”Probability of Loss on Loan Portfolio”, KMV Technical Report.

[33] Vasicek, O. (1991): ”Limiting Loan Loss Probability Distribution”, KMV Technical Report.

[34] Wilde, T. (2001): ”Probing Granularity”, Risk, 14, 103-106.

Table 1: Canonical parameters and summary statistics in exponential families.

Family	Canonical parameter	Cross-sectional ML	Function $c(f)$	Transform $a(y)$
Bernoulli $\mathcal{B}(1,p)$	$f = \log\left(\frac{p}{1-p}\right)$	$\hat{f}_{n,t} = \log\left(\frac{\bar{y}_{n,t}}{1-\bar{y}_{n,t}}\right)$	$c(f) = -\log(1+\exp f)$	$a(y) = y$
Poisson $\mathcal{P}(\lambda)$	$f = \log \lambda$	$\hat{f}_{n,t} = \log \bar{y}_{n,t}$	$c(f) = -\exp f$	$a(y) = y$
Exponential $\gamma(1,\lambda)$	$f = \lambda$	$\hat{f}_{n,t} = 1/\bar{y}_{n,t}$	$c(f) = \log f$	$a(y) = -y$
Gaussian $N(m,1)$	$f = m$	$\hat{f}_{n,t} = \bar{y}_{n,t}$	$c(f) = -f^2/2$	$a(y) = y$
Gaussian $N(0,\sigma^2)$	$f = 1/\sigma^2$	$\hat{f}_{n,t} = 1/\hat{\sigma}^2_{n,t}$	$c(f) = \frac{1}{2}\log f$	$a(y) = -\frac{1}{2}y^2$

In the third column, $\bar{y}_{n,t} = \frac{1}{n}\sum_{i=1}^{n} y_{i,t}$ and $\hat{\sigma}^2_{n,t} = \frac{1}{n}\sum_{i=1}^{n} y^2_{i,t}$ denote the cross-sectional mean and second-order moment, respectively, at date t.

Table 2: Reduced form and structural parameters.

Reduced form parameters			Structural parameters		
$ELGD$	PD	ρ	μ	η	σ
		0.12	4.799	0.766	2.074
0.45	1.5%	0.24	4.799	1.083	1.928
		0.50	4.799	1.564	1.564
		0.12	3.050	0.642	1.739
0.45	5%	0.24	3.050	0.908	1.616
		0.50	3.050	1.311	1.311
		0.12	16.993	2.713	7.346
0.75	1.5%	0.24	16.993	3.836	6.827
		0.50	16.993	5.537	5.537
		0.12	10.669	2.247	6.085
0.75	5%	0.24	10.669	3.178	5.655
		0.50	10.669	4.587	4.587

Table 3: ACF and cross ACF of CSA VaR and GA VaR.

l	$Corr(X_t, X_{t-l})$	$Corr(Y_t, Y_{t-l})$	$Corr(X_t, Y_{t-l})$	$Corr(Y_t, X_{t-l})$
0	1	1	0.86	0.86
1	0.37	0.55	0.34	0.63
2	0.18	0.24	0.17	0.27
3	0.09	0.12	0.08	0.13
4	0.04	0.06	0.04	0.07
5	0.02	0.03	0.02	0.03
6	0.01	0.02	0.01	0.02
7	0.01	0.01	0.01	0.01
8	0.01	0.01	0.00	0.01
9	0.00	0.01	0.00	0.01
10	0.00	0.01	0.00	0.01
11	0.00	0.01	0.00	0.00
12	0.00	0.00	0.00	0.00

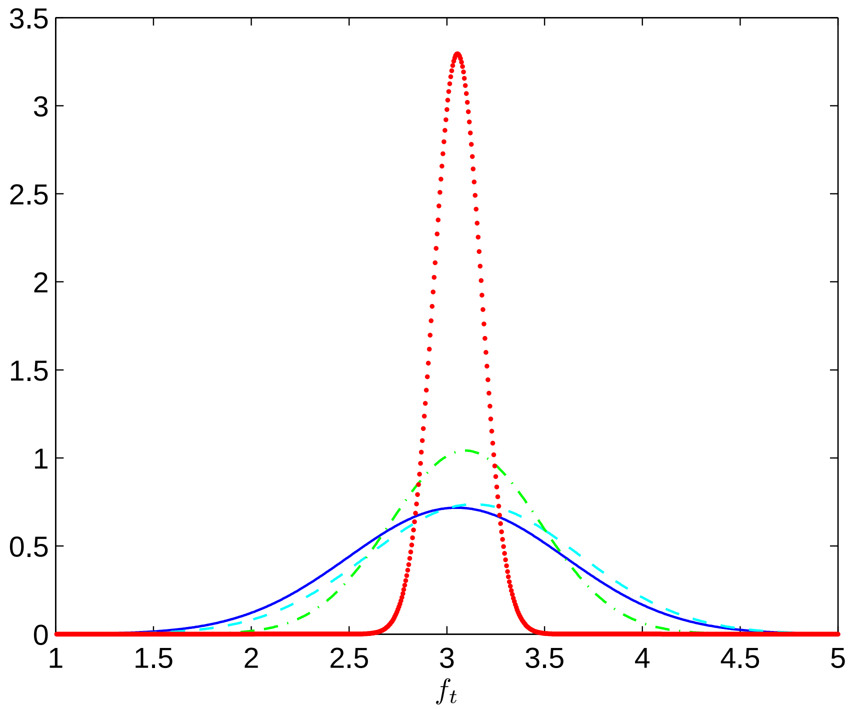
The series are $X_t = VaR_{\infty,t}(\alpha)$ and $Y_t = VaR_{\infty,t}(\alpha) + \frac{1}{n}[GA_{risk}(\alpha) + GA_{filt}(\alpha)]$. The portfolio size is $n = 100$ and the confidence level is $\alpha = 0.995$. The structural parameters are such that $ELGD = 0.45$, $PD = 5\%$, $\rho = 0.12$ and $\gamma = 0.5$ (see Table 2). Correlations are computed by Monte-Carlo simulation on time series of length $T = 100000$.

Table 4: Backtesting of CSA VaR and GA VaR.

	CSA	GA
$E\left[H_t\right]$	0.008	-0.001
$Corr\left(H_t, H_{t-1}\right)$	-0.007	-0.004
$Corr\left(H_t, H_{t-2}\right)$	0.002	-0.000
$Corr\left(H_t, \hat{f}_{n,t-1}\right)$	0.054	-0.022
$Corr\left(H_t, \hat{f}_{n,t-2}\right)$	0.005	0.002
$Corr\left(H_t, W_{n,t-1}/n\right)$	-0.034	0.019
$Corr\left(H_t, W_{n,t-2}/n\right)$	-0.002	0.002

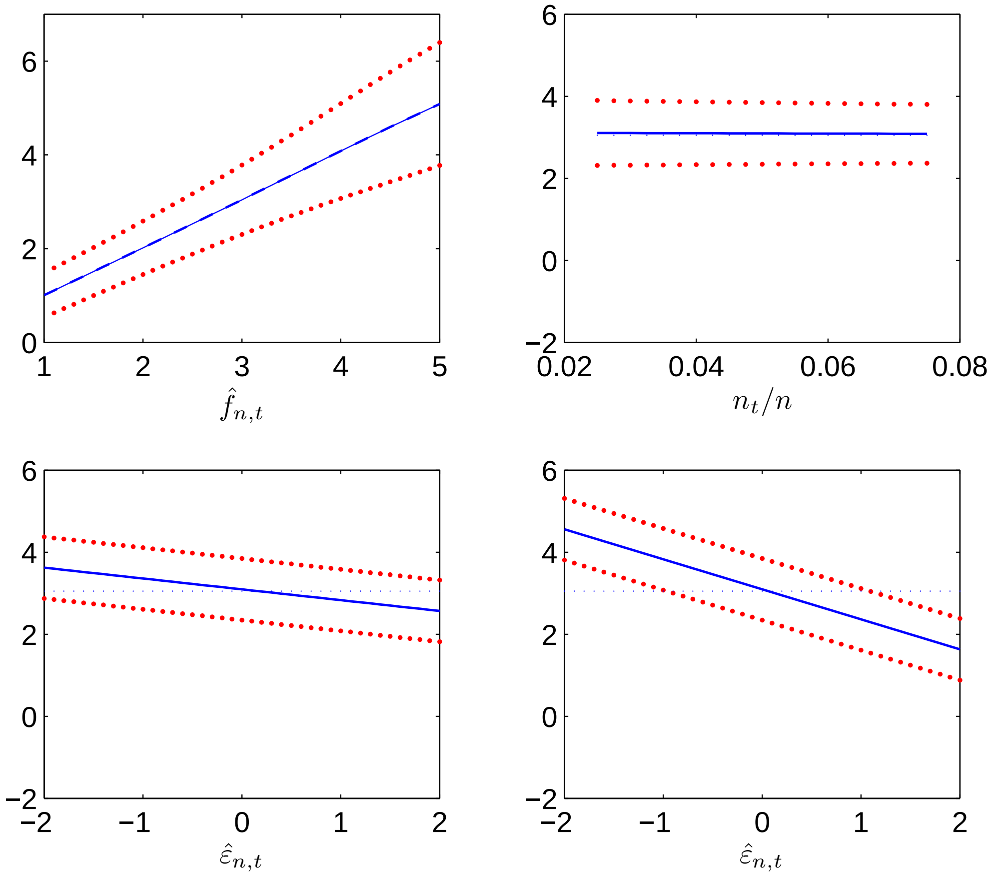
The indicator $H_t = \mathbf{1}_{W_{n,t}/n \geq VaR_{n,t-1}(\alpha)} - (1-\alpha)$ is computed by using $VaR_{n,t-1}(\alpha) = VaR_{\infty,t-1}(\alpha)$ for the CSA VaR and $VaR_{n,t-1}(\alpha) = VaR_{\infty,t-1}(\alpha) + \frac{1}{n}[GA_{risk}(\alpha) + GA_{filt}(\alpha)]$ for the GA VaR. The confidence level is $\alpha = 0.995$. The structural parameters are such that $ELGD = 0.45$, $PD = 5\%$, $\rho = 0.12$ and $\gamma = 0.5$ (see Table 2). All quantities are computed by Monte-Carlo simulation on a time series of length $T = 100000$.

Figure 1: Conditional distribution of F_t given F_{t-1} and approximate filtering distribution of F_t given the micro-information.



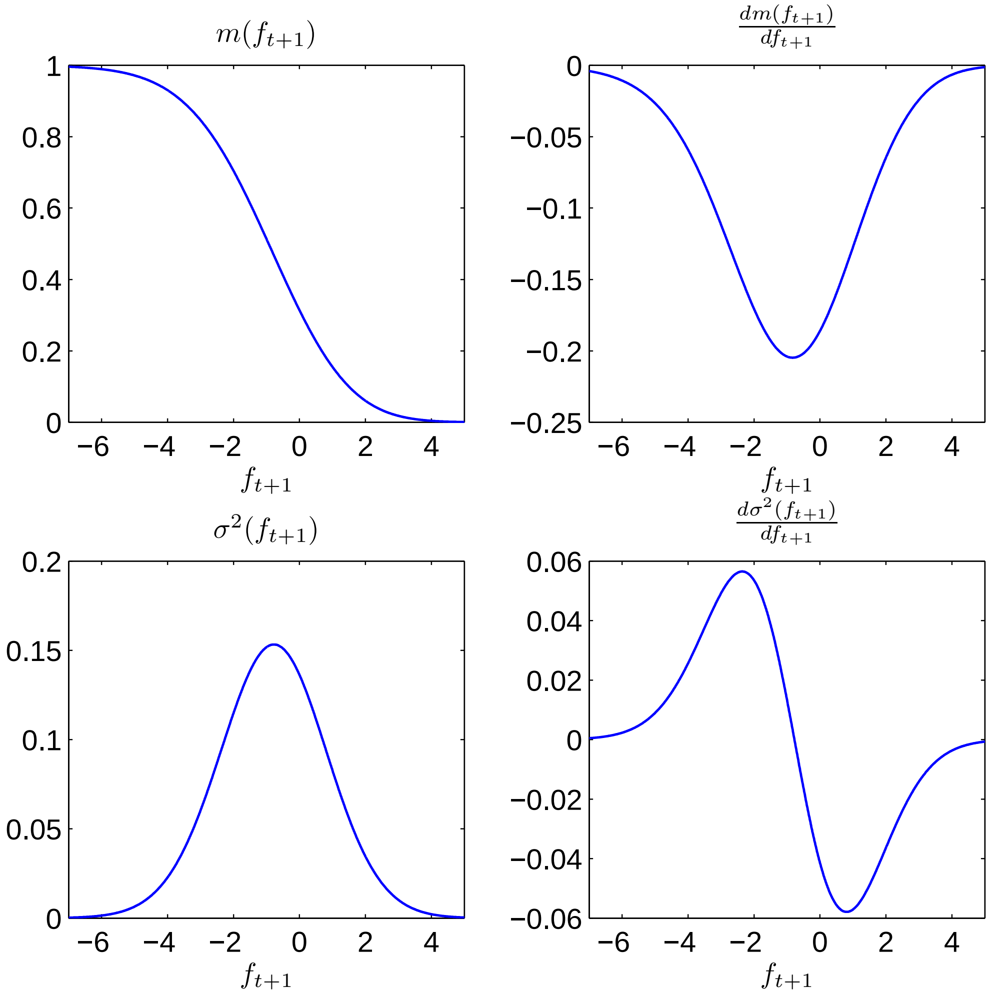
The Figure displays the conditional distribution of F_t given $F_{t-1} = \mu = 3.05$ (solid line) and the approximate filtering distribution of F_t for different values of cross-sectional dimension n, that are $n = 50$ (dashed line), $n = 100$ (dashed-dotted line) and $n = 1000$ (dotted line). The micro-information is such that $\hat{f}_{n,t} = \hat{f}_{n,t-1} = \mu$ and $n_t/n = PD$, for all n. The structural parameters are such that $ELGD = 0.45$, $PD = 5\%$, $\rho = 0.12$ and $\gamma = 0.5$ (see Table 2).

Figure 2: The effect of micro-information on the approximate filtering distribution of F_t.



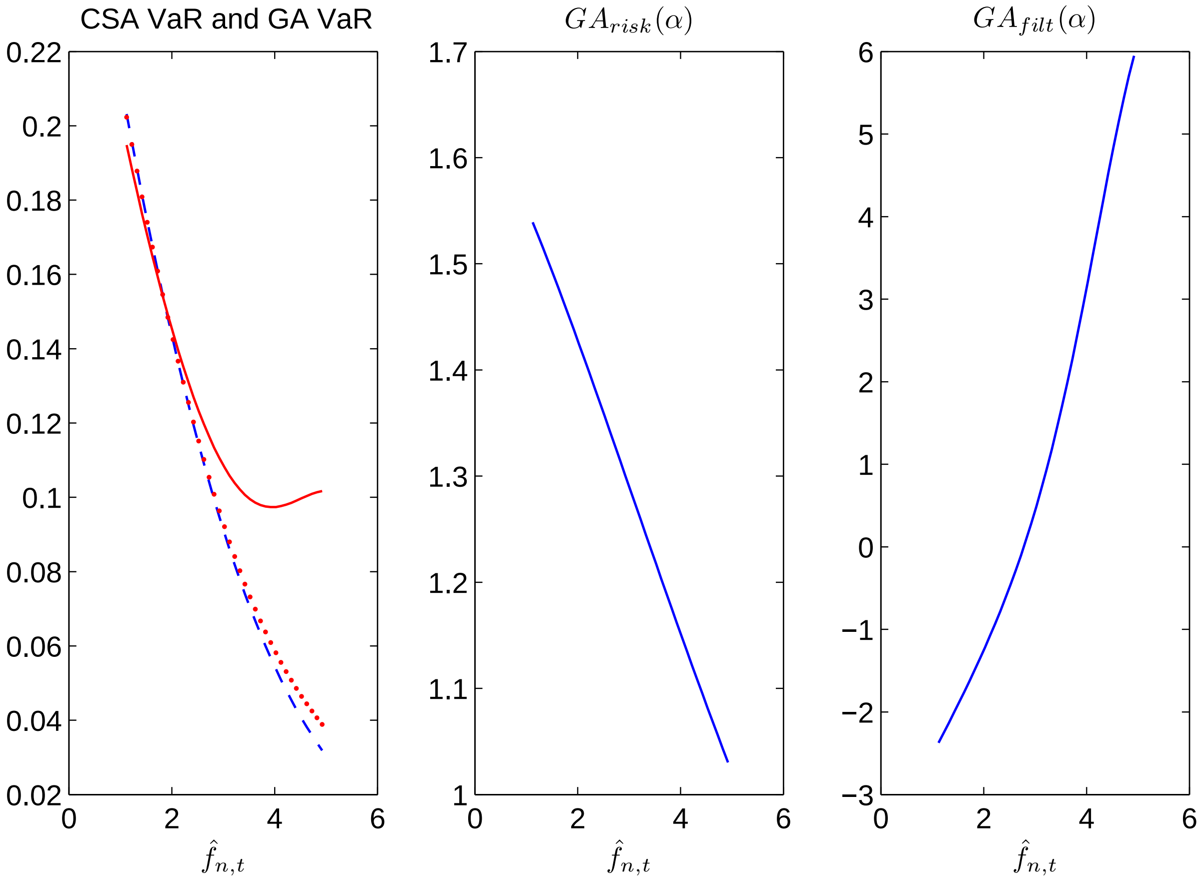
The Figure displays the mean of the approximate filtering distribution of F_t (solid lines), and the 2.5% and 97.5% quantiles of the approximate filtering distribution of F_t (dotted lines), as a function of different micro-information sets for $n = 100$. In the upper left Panel, we set $n_t/n = PD$ and $\hat{\varepsilon}_{n,t} = 0$ and let $\hat{f}_{n,t}$ vary. The structural parameters are such that $ELGD = 0.45$, $PD = 5\%$, $\rho = 0.12$ and $\gamma = 0.5$ (see Table 2). In the upper right Panel, we set $\hat{f}_{n,t} = \mu = 3.05$ and $\hat{\varepsilon}_{n,t} = 0$ and let n_t/n vary. In the lower left Panel, we set $\hat{f}_{n,t} = \mu$ and $n_t/n = PD$ and let $\hat{\varepsilon}_{n,t}$ vary. Finally, in the lower right Panel the same situation is displayed as in the lower left Panel but with $\gamma = 0.95$.

Figure 3: Functions $m(f_{t+1})$ and $\sigma^2(f_{t+1})$ and their first-order derivatives.



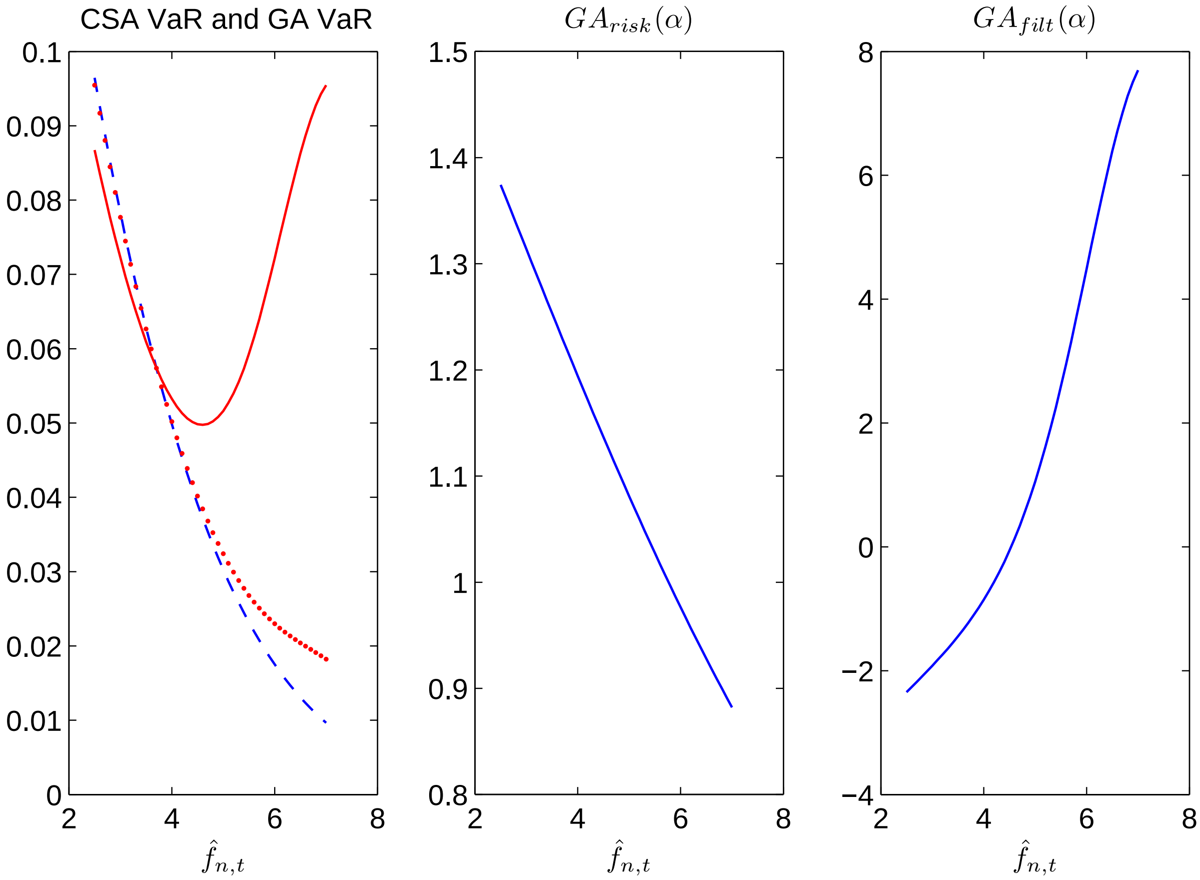
The four Panels display the patterns of functions $m(f_{t+1})$, $dm(f_{t+1})/df_{t+1}$, $\sigma^2(f_{t+1})$, and $d\sigma^2(f_{t+1})/df_{t+1}$, respectively. The structural parameter σ is such that $ELGD = 0.45$, $PD = 5\%$, $\rho = 0.12$ (see Table 2).

Figure 4: CSA and GA VaR as a function of the cross-sectional factor approximation, $PD = 5\%$.



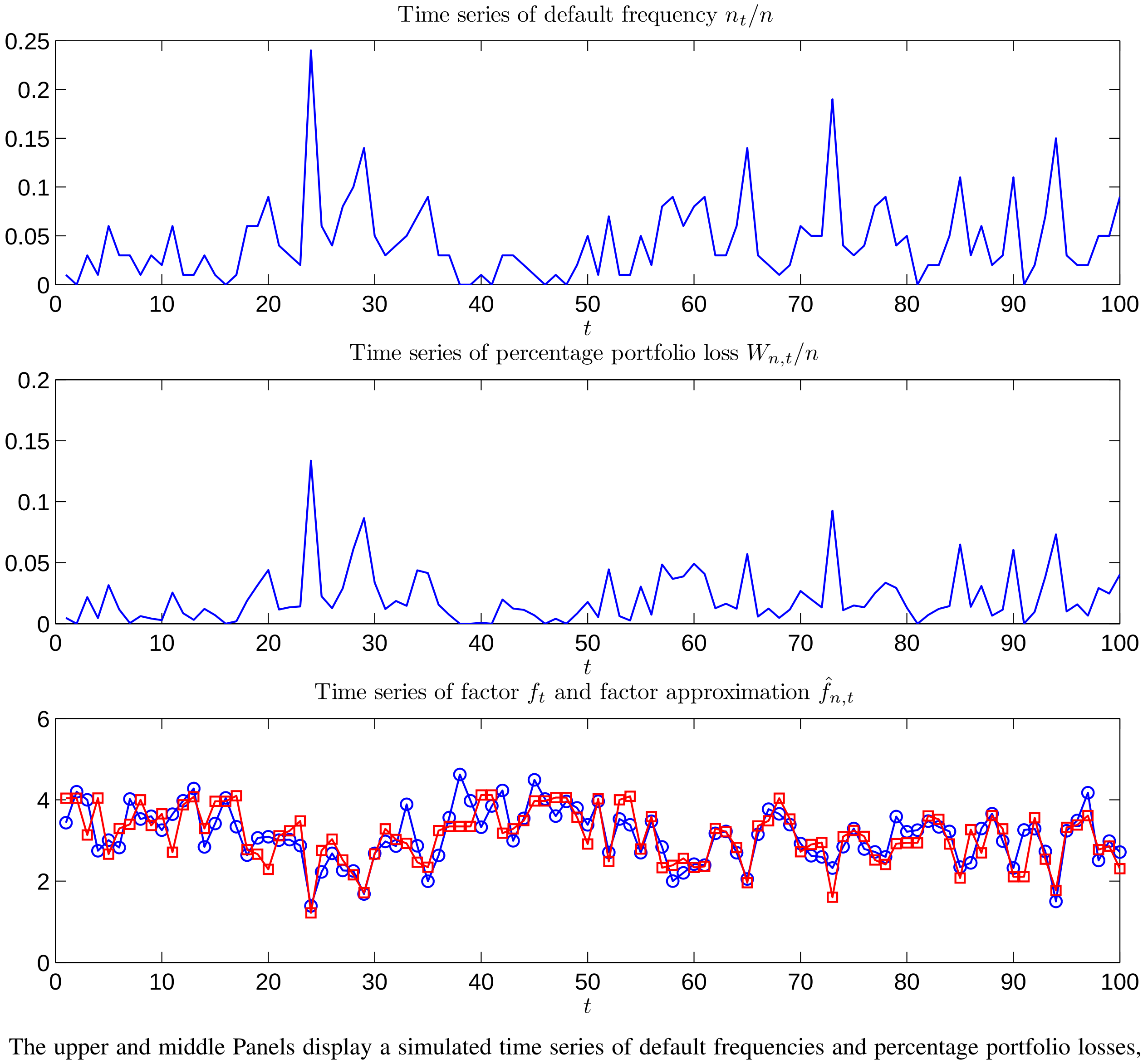
The left Panel displays the CSA VaR (dashed line), the GA VaR for $n = 100$ (solid line) and the GA VaR for $n = 1000$ (dotted line) as functions of the cross-sectional factor approximation $\hat{f}_{n,t}$. The middle and right Panels display the GA component for risk, and the GA component for filtering, respectively. The information set is such that $n_t/n = PD$ and $\hat{f}_{n,t-1} = \mu$. The confidence level is $\alpha = 0.995$. The structural parameters are such that $ELGD = 0.45$, $PD = 5\%$, $\rho = 0.12$ and $\gamma = 0.5$ (see Table 2). In particular, the unconditional factor mean is $\mu = 3.05$.

Figure 5: CSA and GA VaR as a function of the cross-sectional factor approximation, $PD = 1.5\%$.



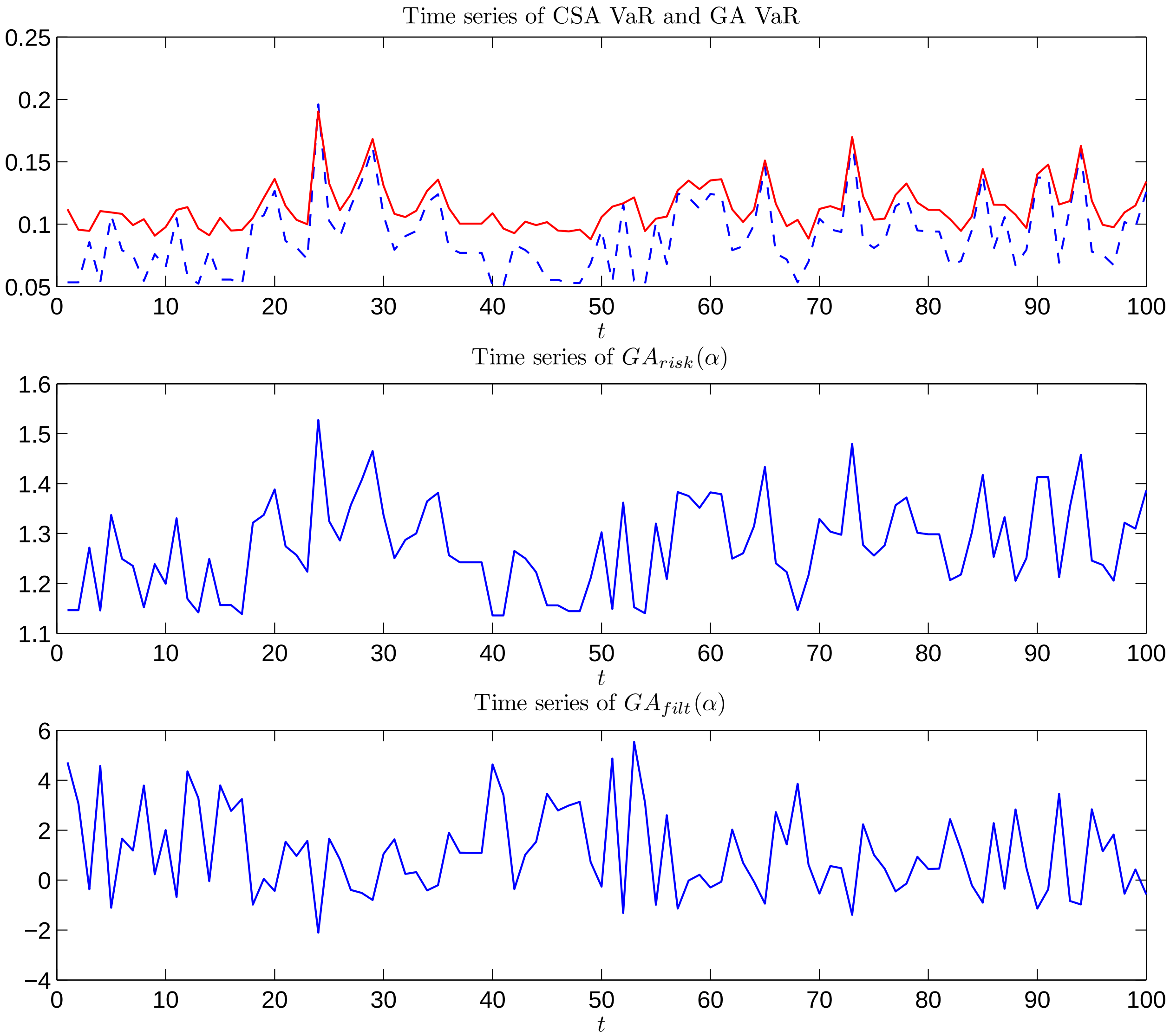
The left Panel displays the CSA VaR (dashed line), the GA VaR for $n = 100$ (solid line) and the GA VaR for $n = 1000$ (dotted line) as functions of the cross-sectional factor approximation $\hat{f}_{n,t}$. The middle and right Panels display the GA component for risk, and the GA component for filtering, respectively. The information set is such that $n_t/n = PD$ and $\hat{f}_{n,t-1} = \mu$. The confidence level is $\alpha = 0.995$. The structural parameters are such that $ELGD = 0.45$, $PD = 1.5\%$, $\rho = 0.12$ and $\gamma = 0.5$ (see Table 2). In particular, the unconditional factor mean is $\mu = 4.799$.

Figure 6: 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 $\gamma = 0.5$ (see Table 2). In particular, the unconditional factor mean is $\mu = 3.05$.

Figure 7: 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 $\alpha = 0.995$. 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 2).

APPENDIX 1: Proof of Proposition 1

(i) Let us first derive the conditional distribution of F_t given $\mathbf{Y}_t, \mathbf{F}_{t-1}, X$. Its density is:

$$p(f_t|\mathbf{Y}_t, \mathbf{F}_{t-1}, X) = \frac{\prod_{i=1}^{n} h_{i,t}(y_{i,t}|f_t) g(f_t|f_{t-1})}{\int \prod_{i=1}^{n} h_{i,t}(y_{i,t}|f_t) g(f_t|f_{t-1}) df_t}.$$

To approximate this distribution at order $1/n$, we consider its Laplace transform:

$$\begin{aligned} E\left[\exp(uF_t)|\mathbf{Y}_t, \mathbf{F}_{t-1}, X\right] &= \frac{\int e^{uf_t} \prod_{i=1}^{n} h_{i,t}(y_{i,t}|f_t) g(f_t|f_{t-1}) df_t}{\int \prod_{i=1}^{n} h_{i,t}(y_{i,t}|f_t) g(f_t|f_{t-1}) df_t} \\ &= \frac{\int \exp\left(uf_t + \sum_{i=1}^{n} \log h_{i,t}(y_{i,t}|f_t) + \log g(f_t|f_{t-1})\right) df_t}{\int \exp\left(\sum_{i=1}^{n} \log h_{i,t}(y_{i,t}|f_t) + \log g(f_t|f_{t-1})\right) df_t}, \ u \in \mathbb{R}, \end{aligned}$$

and perform a Laplace approximation of the integrals in the numerator and denominator for large n. By the same arguments as in the proof of Theorem 1 in Gagliardini, Gouriéroux (2009), we get:

$$\begin{aligned} E\left[\exp(uF_t)|\mathbf{Y}_t, \mathbf{F}_{t-1}, X\right] &= \exp\Bigg[u\left(\hat{f}_{n,t} + \frac{1}{n}\left[I_{n,t}^{-1}\frac{\partial \log g}{\partial f_t}(\hat{f}_{n,t}|f_{t-1}) + \frac{1}{2}I_{n,t}^{-2}K_{n,t}^{(3)}\right]\right) \\ &\quad + \frac{u^2}{2n}I_{n,t}^{-1} + o(1/n)\Bigg]. \end{aligned} \tag{A.1}$$

Since at order $1/n$ the log of $E\left[\exp(uF_t)|\mathbf{Y}_t, \mathbf{F}_{t-1}, X\right]$ involves terms in u and u^2 only, the distribution of F_t given $\mathbf{Y}_t, \mathbf{F}_{t-1}, X$ is Gaussian at order $1/n$:

$$N\left(\hat{f}_{n,t} + \frac{1}{n}\left[I_{n,t}^{-1}\frac{\partial \log g}{\partial f_t}(\hat{f}_{n,t}|f_{t-1}) + \frac{1}{2}I_{n,t}^{-2}K_{n,t}^{(3)}\right], \frac{1}{n}I_{n,t}^{-1}\right). \tag{A.2}$$

(ii) Since $\hat{f}_{n,t-1}$ converges to f_{t-1} as $n \to \infty$, at order $1/n$ we can replace f_{t-1} by $\hat{f}_{n,t-1}$ in the RHS of (A.1) and in (A.2). Thus, the distribution in (A.2) becomes independent of $\mathbf{F}_{t-1}$ up to $o(1/n)$, and coincides with the conditional distribution of F_t given $\mathbf{Y}_t, X$ at order $1/n$. The conclusion follows.

APPENDIX 2: Proof of Proposition 2

Let us expand the integrand in (2.4) around $f_{t+1} = \tilde{f}_{n,t+1}$. We have:

$$\begin{aligned}
\Psi\left(\tilde{y}_{t+1}|y_t, f_t, X\right) &= \int \exp\left[\sum_{i=1}^{n} \log h_{i,t+1}\left(\tilde{y}_{i,t+1}|\tilde{f}_{n,t+1}\right) - \frac{n}{2}\tilde{I}_{n,t+1}\left(f_{t+1} - \tilde{f}_{n,t+1}\right)^2\right.\\
&\quad + \frac{n}{6}\tilde{K}^{(3)}_{n,t+1}\left(f_{t+1} - \tilde{f}_{n,t+1}\right)^3 + \frac{n}{24}\tilde{K}^{(4)}_{n,t+1}\left(f_{t+1} - \tilde{f}_{n,t+1}\right)^4 + \cdots\\
&\quad + \log g\left(\tilde{f}_{n,t+1}|f_t\right) + \frac{\partial \log g}{\partial f_{t+1}}\left(\tilde{f}_{n,t+1}|f_t\right)\left(f_{t+1} - \tilde{f}_{n,t+1}\right)\\
&\quad \left. + \frac{1}{2}\frac{\partial^2 \log g}{\partial f_{t+1}^2}\left(\tilde{f}_{n,t+1}|f_t\right)\left(f_{t+1} - \tilde{f}_{n,t+1}\right)^2 + \cdots\right] df_{t+1}.
\end{aligned}$$

Let us introduce the change of variable:

$$Z^* = \sqrt{n}\tilde{I}^{1/2}_{n,t+1}\left(f_{t+1} - \tilde{f}_{n,t+1}\right) \Leftrightarrow f_{t+1} = \tilde{f}_{n,t+1} + \frac{1}{\sqrt{n}}\tilde{I}^{-1/2}_{n,t+1}Z^*.$$

Then, we get:

$$\begin{aligned}
\Psi\left(\tilde{y}_{t+1}|y_t, f_t, \mathbf{X}\right) &= \prod_{i=1}^{n} h_{i,t+1}\left(\tilde{y}_{i,t+1}|\tilde{f}_{n,t+1}\right) g\left(\tilde{f}_{n,t+1}|f_t\right)\sqrt{\frac{2\pi}{n\tilde{I}_{n,t+1}}}\\
&\quad \cdot E\left\{\exp\left[\frac{1}{\sqrt{n}}\left(\frac{1}{6}\tilde{K}^{(3)}_{n,t+1}\tilde{I}^{-3/2}_{n,t+1}(Z^*)^3 + \tilde{I}^{-1/2}_{n,t+1}\frac{\partial \log g}{\partial f_{t+1}}\left(\tilde{f}_{n,t+1}|f_t\right)Z^*\right)\right.\right.\\
&\quad \left.\left. + \frac{1}{n}\left(\frac{1}{24}\tilde{K}^{(4)}_{n,t+1}\tilde{I}^{-2}_{n,t+1}(Z^*)^4 + \frac{1}{2}\tilde{I}^{-1}_{n,t+1}\frac{\partial^2 \log g}{\partial f_{t+1}^2}\left(\tilde{f}_{n,t+1}|f_t\right)(Z^*)^2\right) + o(1/n)\right]\right\}\\
&=: \prod_{i=1}^{n} h_{i,t+1}\left(\tilde{y}_{i,t+1}|\tilde{f}_{n,t+1}\right) g\left(\tilde{f}_{n,t+1}|f_t\right)\sqrt{\frac{2\pi}{n\tilde{I}_{n,t+1}}}J_n, \text{ say,}
\end{aligned}$$

where the expectation in term J_n is w.r.t. the standard Gaussian variable Z^*. By expanding the exponential function, we get:

$$\begin{aligned}
J_n &= \exp\left\{\frac{1}{n}E\left[\frac{1}{24}\tilde{K}^{(4)}_{n,t+1}\tilde{I}^{-2}_{n,t+1}(Z^*)^4 + \frac{1}{2}\tilde{I}^{-1}_{n,t+1}\frac{\partial^2 \log g}{\partial f_{t+1}^2}\left(\tilde{f}_{n,t+1}|f_t\right)(Z^*)^2\right]\right.\\
&\quad \left. + \frac{1}{2n}E\left[\left(\frac{1}{6}\tilde{K}^{(3)}_{n,t+1}\tilde{I}^{-3/2}_{n,t+1}(Z^*)^3 + \tilde{I}^{-1/2}_{n,t+1}\frac{\partial \log g}{\partial f_{t+1}}\left(\tilde{f}_{n,t+1}|f_t\right)Z^*\right)^2\right] + o(1/n)\right\}\\
&= \exp\left\{\frac{1}{n}\left[\frac{1}{8}\tilde{K}^{(4)}_{n,t+1}\tilde{I}^{-2}_{n,t+1} + \frac{1}{2}\tilde{I}^{-1}_{n,t+1}\frac{\partial^2 \log g}{\partial f_{t+1}^2}\left(\tilde{f}_{n,t+1}|f_t\right) + \frac{5}{24}\left[\tilde{K}^{(3)}_{n,t+1}\right]^2\tilde{I}^{-3}_{n,t+1}\right.\right.\\
&\quad \left.\left. + \frac{1}{2}\tilde{I}^{-1}_{n,t+1}\left(\frac{\partial \log g}{\partial f_{t+1}}\left(\tilde{f}_{n,t+1}|f_t\right)\right)^2 + \frac{1}{2}\tilde{K}^{(3)}_{n,t+1}\tilde{I}^{-2}_{n,t+1}\frac{\partial \log g}{\partial f_{t+1}}\left(\tilde{f}_{n,t+1}|f_t\right)\right] + o(1/n)\right\},
\end{aligned}$$

where we used $E\left[(Z^*)^2\right] = 1$, $E\left[(Z^*)^4\right] = 3$, $E\left[(Z^*)^6\right] = 15$ and that odd-order moments of Z^* vanish. The conclusion follows.

APPENDIX 3: Proof of Proposition 3

The conditional density of y_{t+1} given $\mathbf{Y}_t$ and X is given by:

$$\Psi\left(\tilde{y}_{t+1}|\mathbf{Y}_t, X\right) = \int \Psi\left(\tilde{y}_{t+1}|y_t, f_t, X\right)\Psi\left(f_t|\mathbf{Y}_t, X\right) df_t,$$

where $\Psi\left(\tilde{y}_{t+1}|y_t, f_t, X\right)$ is given in Proposition 2 and $\Psi\left(f_t|\mathbf{Y}_t, X\right)$ is the Gaussian pdf given in Proposition 1 at order $1/n$. Thus, we get:

$$\begin{aligned}
\Psi\left(\tilde{y}_{t+1}|\mathbf{Y}_t, X\right) &= \sqrt{\frac{2\pi}{n\tilde{I}_{n,t+1}}}\prod_{i=1}^{n} h_{i,t+1}\left(\tilde{y}_{i,t+1}|\tilde{f}_{n,t+1}\right) \\
&\cdot \exp\left\{\frac{1}{n}\left[\frac{1}{8}\tilde{K}^{(4)}_{n,t+1}\tilde{I}^{-2}_{n,t+1} + \frac{5}{24}\left[\tilde{K}^{(3)}_{n,t+1}\right]^2\tilde{I}^{-3}_{n,t+1}\right] + o(1/n)\right\} \\
&\cdot \int g\left(\tilde{f}_{n,t+1}|f_t\right)\exp\left\{\frac{1}{2n}\left[\tilde{I}^{-1}_{n,t+1}\left(\frac{\partial^2 \log g\left(\tilde{f}_{n,t+1}|f_t\right)}{\partial f_{t+1}^2} + \left(\frac{\partial \log g\left(\tilde{f}_{n,t+1}|f_t\right)}{\partial f_{t+1}}\right)^2\right)\right.\right. \\
&\left.\left. + \tilde{I}^{-2}_{n,t+1}\tilde{K}^{(3)}_{n,t+1}\frac{\partial \log g\left(\tilde{f}_{n,t+1}|f_t\right)}{\partial f_{t+1}}\right\}\frac{1}{\sqrt{2\pi I^{-1}_{n,t}/n}}\exp\left\{-\frac{nI_{n,t}}{2}\left(f_t - \hat{f}_{n,t} - \frac{1}{n}\xi_{n,t}\right)^2\right\} df_t,
\end{aligned}$$

where:

$$\xi_{n,t} = I^{-1}_{n,t}\frac{\partial \log g\left(\hat{f}_{n,t}|\hat{f}_{n,t-1}\right)}{\partial f_t} + \frac{1}{2}I^{-2}_{n,t}K^{(3)}_{n,t}.$$

The integral:

$$\begin{aligned}
A := \int g\left(\tilde{f}_{n,t+1}|f_t\right)\exp\left\{\frac{1}{2n}\left[\tilde{I}^{-1}_{n,t+1}\left(\frac{\partial^2 \log g\left(\tilde{f}_{n,t+1}|f_t\right)}{\partial f_{t+1}^2} + \left(\frac{\partial \log g\left(\tilde{f}_{n,t+1}|f_t\right)}{\partial f_{t+1}}\right)^2\right)\right.\right. \\
\left.\left. + \tilde{I}^{-2}_{n,t+1}\tilde{K}^{(3)}_{n,t+1}\frac{\partial \log g\left(\tilde{f}_{n,t+1}|f_t\right)}{\partial f_{t+1}}\right\}\frac{1}{\sqrt{2\pi I^{-1}_{n,t}/n}}\exp\left\{-\frac{nI_{n,t}}{2}\left(f_t - \hat{f}_{n,t} - \frac{1}{n}\xi_{n,t}\right)^2\right\} df_t,
\end{aligned}$$

is approximated at order $1/n$ by a Laplace approximation. We expand the integrand around $f_t = \hat{f}_{n,t}$ such that:

$$\begin{aligned}
& g\left(\tilde{f}_{n,t+1}|f_t\right)\exp\left\{-\frac{nI_{n,t}}{2}\left(f_t - \hat{f}_{n,t} - \frac{1}{n}\xi_{n,t}\right)^2\right\} \\
= \;& \exp\left\{\log g\left(\tilde{f}_{n,t+1}|\hat{f}_{n,t}\right) + \frac{\partial \log g\left(\tilde{f}_{n,t+1}|\hat{f}_{n,t}\right)}{\partial f_t}\left(f_t - \hat{f}_{n,t}\right)\right. \\
& + \frac{1}{2}\frac{\partial^2 \log g\left(\tilde{f}_{n,t+1}|\hat{f}_{n,t}\right)}{\partial f_t^2}\left(f_t - \hat{f}_{n,t}\right)^2 + \cdots \\
& \left. - \frac{nI_{n,t}}{2}\left(f_t - \hat{f}_{n,t}\right)^2 + I_{n,t}\xi_{n,t}\left(f_t - \hat{f}_{n,t}\right) - \frac{I_{n,t}}{2n}\xi^2_{n,t}\right\}.
\end{aligned}$$

Then, we introduce the change of variables:

$$Z^* = \sqrt{n} I_{n,t}^{1/2} \left(f_t - \hat{f}_{n,t} \right) \;\Leftrightarrow\; f_t = \hat{f}_{n,t} + \frac{1}{\sqrt{n}} I_{n,t}^{-1/2} Z^*.$$

We get:

$$\begin{aligned}
A \;=\; & g\left(\tilde{f}_{n,t+1}|\hat{f}_{n,t}\right) \exp\left\{ \frac{1}{2n} \left[\tilde{I}_{n,t+1}^{-1} \left(\frac{\partial^2 \log g\left(\tilde{f}_{n,t+1}|\hat{f}_{n,t}\right)}{\partial f_{t+1}^2} + \left(\frac{\partial \log g\left(\tilde{f}_{n,t+1}|\hat{f}_{n,t}\right)}{\partial f_{t+1}} \right)^2 \right) \right.\right. \\
& \left.\left. + \tilde{I}_{n,t+1}^{-2} \tilde{K}_{n,t+1}^{(3)} \frac{\partial \log g\left(\tilde{f}_{n,t+1}|\hat{f}_{n,t}\right)}{\partial f_{t+1}} - I_{n,t}\xi_{n,t}^2 \right] + o(1/n) \right\} \\
& \cdot E\left[\exp\left(\frac{1}{\sqrt{n}} \left[I_{n,t}^{-1/2} \frac{\partial \log g\left(\tilde{f}_{n,t+1}|\hat{f}_{n,t}\right)}{\partial f_t} + I_{n,t}^{1/2}\xi_{n,t} \right] Z^* + \frac{I_{n,t}^{-1}}{2n} \frac{\partial^2 \log g\left(\tilde{f}_{n,t+1}|\hat{f}_{n,t}\right)}{\partial f_t^2} (Z^*)^2 \right) \right],
\end{aligned}$$

where Z^* is a standard Gaussian variable. By developing the exponential function, we have:

$$\begin{aligned}
& E\left[\exp\left(\frac{1}{\sqrt{n}} \left[I_{n,t}^{-1/2} \frac{\partial \log g\left(\tilde{f}_{n,t+1}|\hat{f}_{n,t}\right)}{\partial f_t} + I_{n,t}^{1/2}\xi_{n,t} \right] Z^* + \frac{I_{n,t}^{-1}}{2n} \frac{\partial^2 \log g\left(\tilde{f}_{n,t+1}|\hat{f}_{n,t}\right)}{\partial f_t^2} (Z^*)^2 \right) \right] \\
= \; & \exp\left\{ \frac{1}{2n} \left[I_{n,t}^{-1} \frac{\partial^2 \log g\left(\tilde{f}_{n,t+1}|\hat{f}_{n,t}\right)}{\partial f_t^2} + \left(I_{n,t}^{-1/2} \frac{\partial \log g\left(\tilde{f}_{n,t+1}|\hat{f}_{n,t}\right)}{\partial f_t} + I_{n,t}^{1/2}\xi_{n,t} \right)^2 \right] + o(1/n) \right\}.
\end{aligned}$$

Thus:

$$\begin{aligned}
\Psi\left(\tilde{y}_{t+1}|\mathbf{Y}_t, X\right) \;=\; & \sqrt{\frac{2\pi}{n\tilde{I}_{n,t+1}}} \prod_{i=1}^n h_{i,t+1}\left(\tilde{y}_{i,t+1}|\tilde{f}_{n,t+1}\right) g\left(\tilde{f}_{n,t+1}|\hat{f}_{n,t}\right) \\
& \cdot \exp\left\{ \frac{1}{n} \left[\frac{1}{8} \tilde{K}_{n,t+1}^{(4)} \tilde{I}_{n,t+1}^{-2} + \frac{5}{24} \left[\tilde{K}_{n,t+1}^{(3)}\right]^2 \tilde{I}_{n,t+1}^{-3} \right.\right. \\
& + \frac{1}{2} \tilde{I}_{n,t+1}^{-1} \left(\frac{\partial^2 \log g\left(\tilde{f}_{n,t+1}|\hat{f}_{n,t}\right)}{\partial f_{t+1}^2} + \left(\frac{\partial \log g\left(\tilde{f}_{n,t+1}|\hat{f}_{n,t}\right)}{\partial f_{t+1}} \right)^2 \right) \\
& + \frac{1}{2} \tilde{I}_{n,t+1}^{-2} \tilde{K}_{n,t+1}^{(3)} \frac{\partial \log g\left(\tilde{f}_{n,t+1}|\hat{f}_{n,t}\right)}{\partial f_{t+1}} - \frac{1}{2} I_{n,t}\xi_{n,t}^2 \\
& \left.\left. + \frac{1}{2} I_{n,t}^{-1} \frac{\partial^2 \log g\left(\tilde{f}_{n,t+1}|\hat{f}_{n,t}\right)}{\partial f_t^2} + \frac{1}{2} \left(I_{n,t}^{-1/2} \frac{\partial \log g\left(\tilde{f}_{n,t+1}|\hat{f}_{n,t}\right)}{\partial f_t} + I_{n,t}^{1/2}\xi_{n,t} \right)^2 \right] \right\}.
\end{aligned}$$

By replacing $\xi_{n,t}$ by its definition, and rearranging terms, the conclusion follows.

APPENDIX 4: Proof of Proposition 4

(i) The first-order derivative of the log-density w.r.t. the factor value is:

$$\frac{\partial \log h_{i,t}\left(y_{i,t}|f_t\right)}{\partial f_t} = \frac{1}{h_{i,t}\left(y_{i,t}|f_t\right)}\frac{\partial h_{i,t}\left(y_{i,t}|f_t\right)}{\partial f_t} = a(y_{i,t}) + \frac{dc\left(f_t\right)}{df}.$$

By using $E\left[\frac{1}{h_{i,t}\left(y_{i,t}|f_t\right)}\frac{\partial h_{i,t}\left(y_{i,t}|f_t\right)}{\partial f_t}|F_t = f_t\right] = 0$, we get:

$$\frac{dc\left(f_t\right)}{df} = -E\left[a(y_{i,t})|F_t = f_t\right],$$

and:

$$\frac{\partial \log h_{i,t}\left(y_{i,t}|f_t\right)}{\partial f_t} = a(y_{i,t}) - E\left[a(y_{i,t})|F_t = f_t\right].$$

(ii) The second-order derivative is:

$$\frac{\partial^2 \log h_{i,t}\left(y_{i,t}|f_t\right)}{\partial f_t^2} = \frac{1}{h_{i,t}\left(y_{i,t}|f_t\right)}\frac{\partial^2 h_{i,t}\left(y_{i,t}|f_t\right)}{\partial f_t^2} - \left(\frac{\partial \log h_{i,t}\left(y_{i,t}|f_t\right)}{\partial f_t}\right)^2 = \frac{d^2 c\left(f_t\right)}{df^2}.$$

By using $E\left[\frac{1}{h_{i,t}\left(y_{i,t}|f_t\right)}\frac{\partial^2 h_{i,t}\left(y_{i,t}|f_t\right)}{\partial f_t^2}|F_t = f_t\right] = 0$, we get:

$$\frac{d^2 c\left(f_t\right)}{df^2} = E\left[\frac{\partial^2 \log h_{i,t}\left(y_{i,t}|f_t\right)}{\partial f_t^2}|F_t = f_t\right] = -E\left[\left(\frac{\partial \log h_{i,t}\left(y_{i,t}|f_t\right)}{\partial f_t}\right)^2 |F_t = f_t\right] = -V\left[a(y_{i,t})|F_t = f_t\right].$$

(iii) The third-order derivative is:

$$\frac{\partial^3 \log h_{i,t}\left(y_{i,t}|f_t\right)}{\partial f_t^3} = \frac{d^3 c\left(f_t\right)}{df^3}.$$

Now, we have:

$$\begin{aligned}\frac{\partial^3 \log h_{i,t}\left(y_{i,t}|f_t\right)}{\partial f_t^3} &= \frac{1}{h_{i,t}\left(y_{i,t}|f_t\right)}\frac{\partial^3 h_{i,t}\left(y_{i,t}|f_t\right)}{\partial f_t^3} - \frac{1}{h_{i,t}\left(y_{i,t}|f_t\right)}\frac{\partial^2 h_{i,t}\left(y_{i,t}|f_t\right)}{\partial f_t^2}\left(\frac{\partial \log h_{i,t}\left(y_{i,t}|f_t\right)}{\partial f_t}\right) \\ &\quad -2\left(\frac{\partial \log h_{i,t}\left(y_{i,t}|f_t\right)}{\partial f_t}\right)\left(\frac{\partial^2 \log h_{i,t}\left(y_{i,t}|f_t\right)}{\partial f_t^2}\right).\end{aligned}$$

By substituting:

$$\frac{1}{h_{i,t}\left(y_{i,t}|f_t\right)}\frac{\partial^2 h_{i,t}\left(y_{i,t}|f_t\right)}{\partial f_t^2} = \frac{\partial^2 \log h_{i,t}\left(y_{i,t}|f_t\right)}{\partial f_t^2} + \left(\frac{\partial \log h_{i,t}\left(y_{i,t}|f_t\right)}{\partial f_t}\right)^2,$$

we get:

$$\begin{aligned}\frac{\partial^3 \log h_{i,t}\left(y_{i,t}|f_t\right)}{\partial f_t^3} &= \frac{1}{h_{i,t}\left(y_{i,t}|f_t\right)}\frac{\partial^3 h_{i,t}\left(y_{i,t}|f_t\right)}{\partial f_t^3} - 3\left(\frac{\partial \log h_{i,t}\left(y_{i,t}|f_t\right)}{\partial f_t}\right)\left(\frac{\partial^2 \log h_{i,t}\left(y_{i,t}|f_t\right)}{\partial f_t^2}\right) \\ &\quad -\left(\frac{\partial \log h_{i,t}\left(y_{i,t}|f_t\right)}{\partial f_t}\right)^3.\end{aligned}$$

By using $E\left[\frac{1}{h_{i,t}\left(y_{i,t}|f_t\right)}\frac{\partial^3 h_{i,t}\left(y_{i,t}|f_t\right)}{\partial f_t^3}|F_t=f_t\right]=0$, $E\left[\frac{\partial \log h_{i,t}\left(y_{i,t}|f_t\right)}{\partial f_t}|F_t=f_t\right]=0$ and $\frac{\partial^2 \log h_{i,t}\left(y_{i,t}|f_t\right)}{\partial f_t^2}=\frac{d^2 c\left(f_t\right)}{df^2}$, we get:

$$\begin{aligned}
\frac{d^3 c\left(f_t\right)}{df^3} &= E\left[\frac{\partial^3 \log h_{i,t}\left(y_{i,t}|f_t\right)}{\partial f_t^3}|F_t=f_t\right]=-E\left[\left(\frac{\partial \log h_{i,t}\left(y_{i,t}|f_t\right)}{\partial f_t}\right)^3|F_t=f_t\right] \\
&= -E\left[\left(a(y_{i,t})-E\left[a(y_{i,t})|f_t\right]\right)^3|F_t=f_t\right].
\end{aligned}$$

(iv) The fourth-order derivative is:

$$\frac{\partial^4 \log h_{i,t}\left(y_{i,t}|f_t\right)}{\partial f_t^4}=\frac{d^4 c\left(f_t\right)}{df^4}.$$

Now, we have:

$$\begin{aligned}
\frac{\partial^4 \log h_{i,t}\left(y_{i,t}|f_t\right)}{\partial f_t^4} &= \frac{1}{h_{i,t}\left(y_{i,t}|f_t\right)}\frac{\partial^4 h_{i,t}\left(y_{i,t}|f_t\right)}{\partial f_t^4}-\frac{1}{h_{i,t}\left(y_{i,t}|f_t\right)}\frac{\partial^3 h_{i,t}\left(y_{i,t}|f_t\right)}{\partial f_t^3}\left(\frac{\partial \log h_{i,t}\left(y_{i,t}|f_t\right)}{\partial f_t}\right) \\
&\quad -3\left(\frac{\partial^2 \log h_{i,t}\left(y_{i,t}|f_t\right)}{\partial f_t^2}\right)^2-3\left(\frac{\partial \log h_{i,t}\left(y_{i,t}|f_t\right)}{\partial f_t}\right)\left(\frac{\partial^3 \log h_{i,t}\left(y_{i,t}|f_t\right)}{\partial f_t^3}\right) \\
&\quad -3\left(\frac{\partial \log h_{i,t}\left(y_{i,t}|f_t\right)}{\partial f_t}\right)^2\frac{\partial^2 \log h_{i,t}\left(y_{i,t}|f_t\right)}{\partial f_t^2}.
\end{aligned}$$

By substituting:

$$\begin{aligned}
\frac{1}{h_{i,t}\left(y_{i,t}|f_t\right)}\frac{\partial^3 h_{i,t}\left(y_{i,t}|f_t\right)}{\partial f_t^3} &= \frac{\partial^3 \log h_{i,t}\left(y_{i,t}|f_t\right)}{\partial f_t^3}+3\left(\frac{\partial \log h_{i,t}\left(y_{i,t}|f_t\right)}{\partial f_t}\right)\left(\frac{\partial^2 \log h_{i,t}\left(y_{i,t}|f_t\right)}{\partial f_t^2}\right) \\
&\quad +\left(\frac{\partial \log h_{i,t}\left(y_{i,t}|f_t\right)}{\partial f_t}\right)^3,
\end{aligned}$$

we get:

$$\begin{aligned}
\frac{\partial^4 \log h_{i,t}\left(y_{i,t}|f_t\right)}{\partial f_t^4} &= \frac{1}{h_{i,t}\left(y_{i,t}|f_t\right)}\frac{\partial^4 h_{i,t}\left(y_{i,t}|f_t\right)}{\partial f_t^4} \\
&\quad -6\left(\frac{\partial \log h_{i,t}\left(y_{i,t}|f_t\right)}{\partial f_t}\right)^2\left(\frac{\partial^2 \log h_{i,t}\left(y_{i,t}|f_t\right)}{\partial f_t^2}\right)-\left(\frac{\partial \log h_{i,t}\left(y_{i,t}|f_t\right)}{\partial f_t}\right)^4 \\
&\quad -3\left(\frac{\partial^2 \log h_{i,t}\left(y_{i,t}|f_t\right)}{\partial f_t^2}\right)^2-4\left(\frac{\partial \log h_{i,t}\left(y_{i,t}|f_t\right)}{\partial f_t}\right)\left(\frac{\partial^3 \log h_{i,t}\left(y_{i,t}|f_t\right)}{\partial f_t^3}\right).
\end{aligned}$$

By using that $E\left[\frac{1}{h_{i,t}\left(y_{i,t}|f_t\right)}\frac{\partial^4 h_{i,t}\left(y_{i,t}|f_t\right)}{\partial f_t^4}|F_t=f_t\right]=0$, $E\left[\frac{\partial \log h_{i,t}\left(y_{i,t}|f_t\right)}{\partial f_t}|F_t=f_t\right]=0$, $\frac{\partial^2 \log h_{i,t}\left(y_{i,t}|f_t\right)}{\partial f_t^2}=\frac{d^2 c\left(f_t\right)}{df^2}=-E\left[\left(\frac{\partial^2 \log h_{i,t}\left(y_{i,t}|f_t\right)}{\partial f_t^2}\right)^2|F_t=f_t\right]$ and $\frac{\partial^3 \log h_{i,t}\left(y_{i,t}|f_t\right)}{\partial f_t^3}=\frac{d^3 c\left(f_t\right)}{df^3}$, we get:

$$\begin{aligned}
\frac{d^4 c\left(f_t\right)}{df^4} &= E\left[\frac{\partial^4 \log h_{i,t}\left(y_{i,t}|f_t\right)}{\partial f_t^4}|F_t=f_t\right] \\
&= 3E\left[\left(\frac{\partial^2 \log h_{i,t}\left(y_{i,t}|f_t\right)}{\partial f_t^2}\right)^2|F_t=f_t\right]-E\left[\left(\frac{\partial \log h_{i,t}\left(y_{i,t}|F_t=f_t\right)}{\partial f_t}\right)^4|F_t=f_t\right] \\
&= -\left\{E\left[\left(a(y_{i,t})-E\left[a(y_{i,t})|F_t=f_t\right]\right)^4|F_t=f_t\right]-3V\left[a(y_{i,t})|F_t=f_t\right]^2\right\}.
\end{aligned}$$

APPENDIX 5: Proof of Proposition 5

Let us first rewrite the RHS of (4.6). By using $\log g(\hat{f}_{n,t}|\hat{f}_{n,t-1};\theta) = -\frac{1}{2}\log(2\pi\sigma^2) - \frac{\hat{\epsilon}_{n,t}(\theta)^2}{2\sigma^2}$, $\frac{\partial \log g(\hat{f}_{n,t}|\hat{f}_{n,t-1};\theta)}{\partial f_t} = -\frac{\hat{\epsilon}_{n,t}(\theta)}{\sigma^2}$, $\frac{\partial^2 \log g(\hat{f}_{n,t}|\hat{f}_{n,t-1};\theta)}{\partial f_t^2} = -\frac{1}{\sigma^2}$, $\frac{\partial \log g(\hat{f}_{n,t}|\hat{f}_{n,t-1};\theta)}{\partial f_{t-1}} = \frac{\gamma\hat{\epsilon}_{n,t}(\theta)}{\sigma^2}$ and $\frac{\partial^2 \log g(\hat{f}_{n,t}|\hat{f}_{n,t-1};\theta)}{\partial f_{t-1}^2} = -\frac{\gamma^2}{\sigma^2}$, where $\hat{\epsilon}_{n,t}(\theta) = \hat{f}_{n,t} - \mu - \gamma(\hat{f}_{n,t-1} - \mu)$ and $\sigma = \eta\sqrt{1-\gamma^2}$, we get:

$$\begin{aligned}
\log p(y_t|\mathbf{Y}_{t-1}, X;\theta) &= -\frac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2n\sigma^2}(I_{n,t}^{-1} + \gamma^2 I_{n,t-1}^{-1}) \\
&\quad -\frac{1}{2\sigma^2}\left[1 - \frac{1}{n\sigma^2}(I_{n,t}^{-1} + \gamma^2 I_{n,t-1}^{-1})\right]\hat{\varepsilon}_{n,t}(\theta)^2 \\
&\quad -\frac{1}{2n\sigma^2}\left(I_{n,t}^{-1}K_{n,t}^{(3)} - \gamma I_{n,t-1}^{-1}K_{n,t-1}^{(3)}\right)\hat{\varepsilon}_{n,t}(\theta) - \frac{\gamma}{n\sigma^4}I_{n,t-1}^{-1}\hat{\varepsilon}_{n,t}(\theta)\hat{\varepsilon}_{n,t-1}(\theta).
\end{aligned}$$

Thus from (4.5), the GA log-likelihood function can be written as:

$$\mathcal{L}_{nT}^{\text{GA}}(\theta) = -\frac{1}{2}\omega_n(\theta) - \frac{1}{2T\sigma^2}U_n(\theta)'\left(Id_T - \frac{1}{n\sigma^2}B_n(\theta)\right)U_n(\theta) + o(1/n), \tag{A.3}$$

up to a constant term in θ, where $U_n(\theta)$ is a $(T,1)$ vector with elements:

$$U_{n,t}(\theta) = \hat{\varepsilon}_{n,t-1}(\theta) + \frac{1}{2n}\left(I_{n,t}^{-1}K_{n,t}^{(3)} - \gamma I_{n,t-1}^{-1}K_{n,t-1}^{(3)}\right) = \xi_{n,t} - \mu - \gamma(\xi_{n,t-1} - \mu),$$

the symmetric (T,T) matrix $B_n(\theta)$ has elements equal to $I_{n,t}^{-1} + \gamma^2 I_{n,t-1}^{-1}$ in position (t,t), $-\gamma I_{n,t-1}^{-1}$ in positions $(t-1,t)$ and $(t,t-1)$, and zeros otherwise, and the scalar $\omega_n(\theta)$ is given by $\omega_n(\theta) = \log(2\pi\sigma^2) + \frac{1}{\sigma^2 nT}\sum_{t=1}^{T}(I_{n,t}^{-1} + \gamma^2 I_{n,t-1}^{-1})$.

Now, we have:

$$\frac{1}{\sigma^2}\left(Id_T - \frac{1}{n\sigma^2}B_n(\theta)\right) = \Omega_n(\theta)^{-1} + o(1/n), \tag{A.4}$$

where $\Omega_n(\theta) = \sigma^2 Id_T + \frac{1}{n}B_n(\theta)$. Moreover:

$$\begin{aligned}
\frac{1}{T}\log\det\Omega_n(\theta) &= \log\sigma^2 + \frac{1}{T}\log\det\left(Id_T + \frac{1}{n\sigma^2}B_n(\theta)\right) \\
&= \log\sigma^2 + \frac{1}{T}\log\left(1 + \frac{1}{n\sigma^2}trB_n(\theta) + o(T/n)\right) \\
&= \log\sigma^2 + \frac{1}{\sigma^2 nT}trB_n(\theta) + o(1/n) = \omega_n(\theta) + o(1/n).
\end{aligned} \tag{A.5}$$

By replacing (A.4) and (A.5) into (A.3), we get:

$$\mathcal{L}_{nT}^{\text{GA}}(\theta) = -\frac{1}{2T}\log\det\Omega_n(\theta) - \frac{1}{2T\sigma^2}U_n(\theta)'\Omega_n(\theta)^{-1}U_n(\theta) + o(1/n).$$

By noting that $\Omega_n(\theta)$ is the variance-covariance matrix of the errors $\sigma\varepsilon_t + \frac{1}{\sqrt{n}}I_{n,t}^{-1/2}u_t - \gamma\frac{1}{\sqrt{n}}I_{n,t-1}^{-1/2}u_{t-1}$, where (ε_t) and (u_t) are independent Gaussian white noise processes, the conclusion follows.

APPENDIX 6: The Value of the Firm model

1. Unconditional ELGD [proof of equation (6.7)]:

We use the next Lemma 1:

Lemma 1: *Let Z be a standard Gaussian variable. Then:*

$$E[\exp(aZ)|Z<b] = \frac{E[\exp(aZ)\mathbb{1}_{Z<b}]}{P[Z<b]} = \exp\left(\frac{a^2}{2}\right)\frac{\Phi(b-a)}{\Phi(b)},$$

for any $a, b \in \mathbb{R}$.

Proof of Lemma 1: We have:

$$E[\exp(aZ)\mathbb{1}_{Z<b}] = \int_{-\infty}^{b} e^{az}\phi(z)dz = \exp\left(\frac{a^2}{2}\right)\int_{-\infty}^{b}\phi(z-a)dz = \exp\left(\frac{a^2}{2}\right)\Phi(b-a).$$

The conclusion follows.

QED

Let us now prove equation (6.7). Since $\log(A_{i,t}/L_{i,t}) = F_t + \sigma u_{i,t} \sim N(\mu, \eta^2+\sigma^2)$, we have:

$$ELGD = 1 - E\left[\exp\left(F_t+\sigma u_{i,t}\right)|F_t+\sigma u_{i,t}<0\right] = 1 - e^{\mu}E\left[\exp\left(\sqrt{\eta^2+\sigma^2}Z\right)|Z < -\frac{\mu}{\sqrt{\eta^2+\sigma^2}}\right],$$

where $Z \sim N(0,1)$. Then, equation (6.7) follows from Lemma 1 with $a = \sqrt{\eta^2+\sigma^2}$ and $b = -\frac{\mu}{\sqrt{\eta^2+\sigma^2}}$.

2. Parameterization in terms of $PD, \rho, ELGD$ [proof of equations (6.8) and (6.9)]

From equations (6.4) and (6.5) we have $\mu = -\tau\Phi^{-1}(PD)$, $\eta = \tau\sqrt{\rho}$ and $\sigma = \tau\sqrt{1-\rho}$ where $\tau = \sqrt{\eta^2+\sigma^2}$, which yields (6.8). Moreover, equation (6.7) can be rewritten as:

$$ELGD\cdot PD = PD - \exp\left(-\tau\Phi^{-1}(PD)+\frac{1}{2}\tau^2\right)\Phi\left[\Phi^{-1}(PD)-\tau\right].$$

Thus, parameter τ solves the equation $A(\tau) = ELGD \cdot PD$, where function A is given by:

$$A(\tau) = PD - \exp\left(-\tau\Phi^{-1}(PD)+\frac{1}{2}\tau^2\right)\Phi\left[\Phi^{-1}(PD)-\tau\right], \quad \tau \geq 0.$$

Let us now prove that function A is monotone. Its derivative is given by:

$$\begin{aligned}\frac{dA(\tau)}{d\tau} &= \left[(\Phi^{-1}(PD)-\tau)\Phi\left(\Phi^{-1}(PD)-\tau\right)+\phi(\Phi^{-1}(PD)-\tau)\right]\exp\left(-\tau\Phi^{-1}(PD)+\frac{1}{2}\tau^2\right)\\ &= B[\Phi^{-1}(PD)-\tau]\exp\left(-\tau\Phi^{-1}(PD)+\frac{1}{2}\tau^2\right),\end{aligned}$$

where $B(x) = x\Phi(x)+\phi(x)$. We have $dB(x)/dx = \Phi(x) \geq 0$, and $B(-\infty) = 0$. Thus $B(x) \geq 0$ for all x. We deduce that $dA(\tau)/\tau \geq 0$ and function A is monotone increasing on $\mathbb{R}^+$, with $A(0) = 0$ and $A(+\infty) = PD$. Thus, equation (6.9) admits a unique solution as long as $0 < ELGD < 1$.

3. Approximate filtering distribution [proof of equation (6.13)]

From (6.10) the cross-sectional log-likelihood at date t is given by:

$$\mathcal{L}_n(f_t) = \sum_{i=1}^{n} \log h(y_{i,t}|f_t) = -\frac{1}{2\sigma^2} \sum_{i:y_{i,t}>0} [\log(1-y_{i,t}) - f_t]^2 + (n-n_t)\log\Phi(f_t/\sigma).$$

The derivatives of the log-likelihood function w.r.t. f_t are:

$$\begin{aligned}
\frac{\partial \mathcal{L}_n(f_t)}{df_t} &= \frac{1}{\sigma^2} \sum_{i:y_{i,t}>0} [\log(1-y_{i,t}) - f_t] + (n-n_t)\frac{1}{\sigma}\lambda(f_t/\sigma), \\
\frac{\partial^2 \mathcal{L}_n(f_t)}{df_t^2} &= -\frac{n_t}{\sigma^2} + (n-n_t)\frac{1}{\sigma^2}\lambda'(f_t/\sigma), \\
\frac{\partial^3 \mathcal{L}_n(f_t)}{df_t^3} &= (n-n_t)\frac{1}{\sigma^3}\lambda''(f_t/\sigma),
\end{aligned}$$

where:

$$\lambda(x) = \frac{\phi(x)}{\Phi(x)}, \quad \lambda'(x) = -\lambda(x)[x+\lambda(x)], \quad \lambda''(x) = -\lambda(x)\{1-[x+\lambda(x)][x+2\lambda(x)]\}.$$

From $K_{n,t}^{(l)} = \frac{1}{n}\frac{\partial^l \mathcal{L}_n(\hat{f}_{n,t})}{df_t^l}$, for $l = 2, 3$, equations (6.14) and (6.15) follow. Finally, the derivative of the log transition density of the factor is:

$$\frac{\partial \log g}{\partial f_t}(f_t|f_{t-1}) = -\frac{f_t - \mu - \gamma(f_{t-1}-\mu)}{\eta^2(1-\gamma^2)},$$

and from Proposition 1 we get (6.13).

4. Function $\sigma^2(f_{t+1})$ and its derivative [proof of equations (6.18) and (6.19)]

We have:

$$\begin{aligned}
m^{(2)}(f_{t+1}) &:= E[y_{i,t+1}^2|F_{t+1}=f_{t+1}] = E[\mathbb{1}_{u_{i,t+1}<-F_{t+1}/\sigma}(1-\exp(F_{t+1}+\sigma u_{i,t+1}))^2|F_{t+1}=f_{t+1}] \\
&= E[\mathbb{1}_{u_{i,t+1}<-F_{t+1}/\sigma}|F_{t+1}=f_{t+1}] - 2\exp(F_{t+1})E[\mathbb{1}_{u_{i,t+1}<-F_{t+1}/\sigma}\exp(\sigma u_{i,t+1})|F_{t+1}=f_{t+1}] \\
&\quad + \exp(2F_{t+1})E[\mathbb{1}_{u_{i,t+1}<-F_{t+1}/\sigma}\exp(2\sigma u_{i,t+1})|F_{t+1}=f_{t+1}].
\end{aligned}$$

Then, from Lemma 1 we get:

$$\begin{aligned}
m^{(2)}(f_{t+1}) &= \Phi\left(-f_{t+1}/\sigma\right) - 2\exp\left(f_{t+1}+\frac{\sigma^2}{2}\right)\Phi\left(-f_{t+1}/\sigma-\sigma\right) \\
&\quad + \exp(2f_{t+1}+2\sigma^2)\Phi\left(-f_{t+1}/\sigma-2\sigma\right) \\
&= m(f_{t+1}) - \exp\left(f_{t+1}+\frac{\sigma^2}{2}\right)\Phi\left(-f_{t+1}/\sigma-\sigma\right) \\
&\quad + \exp(2f_{t+1}+2\sigma^2)\Phi\left(-f_{t+1}/\sigma-2\sigma\right).
\end{aligned}$$

Thus, by using $\sigma^2(f_{t+1}) = m^{(2)}(f_{t+1}) - m(f_{t+1})^2$, equation (6.18) follows.

To compute the derivative of $m^{(2)}(f_{t+1})$ w.r.t. f_{t+1}, write $m^{(2)}(f_{t+1}) = \int_{-\infty}^{-f_{t+1}/\sigma}[1-\exp(f_{t+1}+\sigma u)]^2\phi(u)du$. Then:

$$\begin{aligned}
\frac{dm^{(2)}(f_{t+1})}{df_{t+1}} &= \frac{d}{df_{t+1}}\int_{-\infty}^{-f_{t+1}/\sigma}[1-\exp(f_{t+1}+\sigma u)]^2\phi(u)du \\
&= -2\exp(f_{t+1})\int_{-\infty}^{-f_{t+1}/\sigma}\exp(\sigma u)[1-\exp(f_{t+1}+\sigma u)]\phi(u)du \\
&= -2\exp(f_{t+1})E\left[\mathbb{1}_{u_{i,t+1}<-F_{t+1}/\sigma}\left\{\exp(\sigma u_{i,t+1})-\exp(F_{t+1}+2\sigma u_{i,t+1})\right\}|F_{t+1}=f_{t+1}\right] \\
&= -2\exp\left(f_{t+1}+\frac{\sigma^2}{2}\right)\Phi\left(-f_{t+1}/\sigma-\sigma\right)+2\exp\left(2f_{t+1}+2\sigma^2\right)\Phi\left(-f_{t+1}/\sigma-2\sigma\right).
\end{aligned}$$

Then, we get the derivative of $\sigma^2(f_{t+1})$:

$$\begin{aligned}
&\frac{d\sigma^2(f_{t+1})}{df_{t+1}} \\
=\ &\frac{dm^{(2)}(f_{t+1})}{df_{t+1}} - 2m(f_{t+1})\frac{dm(f_{t+1})}{df_{t+1}} \\
=\ &-2\exp\left(f_{t+1}+\frac{\sigma^2}{2}\right)\Phi\left(-f_{t+1}/\sigma-\sigma\right)+2\exp\left(2f_{t+1}+2\sigma^2\right)\Phi\left(-f_{t+1}/\sigma-2\sigma\right) \\
&+2\left[\Phi(-f_{t+1}/\sigma)-\exp\left(f_{t+1}+\frac{\sigma^2}{2}\right)\Phi\left(-f_{t+1}/\sigma-\sigma\right)\right]\exp\left(f_{t+1}+\frac{\sigma^2}{2}\right)\Phi\left(-f_{t+1}/\sigma-\sigma\right) \\
=\ &-2\exp\left(f_{t+1}+\frac{\sigma^2}{2}\right)\Phi\left(-f_{t+1}/\sigma-\sigma\right)[1-\Phi(-f_{t+1}/\sigma)]+2\exp\left(2f_{t+1}+2\sigma^2\right)\Phi\left(-f_{t+1}/\sigma-2\sigma\right) \\
&-2\exp\left(2f_{t+1}+\sigma^2\right)\left[\Phi\left(-f_{t+1}/\sigma-\sigma\right)\right]^2,
\end{aligned}$$

which yields equation (6.19).

5. Derivation of GA [proof of Proposition 8]

The GA is derived by using Proposition 7. Let us first consider the GA for risk. We have:

$$\frac{d\sigma^2[m^{-1}(w)]}{dw} = \frac{\frac{d\sigma^2}{df_{t+1}}[m^{-1}(w)]}{\frac{dm}{df_{t+1}}[m^{-1}(w)]},$$

and from (6.22):

$$\log f_{\infty,t}(w) = -\frac{[m^{-1}(w)-\mu-\gamma(\hat{f}_{n,t}-\mu)]^2}{2\eta^2(1-\gamma^2)} - m^{-1}(w) - \log\Phi\left[-\frac{m^{-1}(w)}{\sigma}-\sigma\right],$$

up to an additive constant, which yields:

$$\begin{aligned}
\frac{\partial\log f_{\infty,t}(w)}{\partial w} &= -\left[\frac{m^{-1}(w)-\mu-\gamma(\hat{f}_{n,t}-\mu)}{\eta^2(1-\gamma^2)}+1-\frac{1}{\sigma}\lambda\left(-\frac{m^{-1}(w)}{\sigma}-\sigma\right)\right]\frac{dm^{-1}(w)}{dw} \\
&= \left[-\frac{m^{-1}(w)-\mu-\gamma(\hat{f}_{n,t}-\mu)}{\eta^2(1-\gamma^2)}-1+\frac{1}{\sigma}\lambda\left(-\frac{m^{-1}(w)}{\sigma}-\sigma\right)\right]\frac{1}{\frac{dm}{df_{t+1}}[m^{-1}(w)]}.
\end{aligned}$$

Then, from Proposition 7 and equations (6.23) and (6.24), the GA for risk (6.25) follows.

Let us now consider the GA for filtering. From (6.21), the first- and second-order derivatives of function $a(w, f_t, 0)$ w.r.t. f_t at $\hat{f}_{n,t}$ are given by:

$$\begin{aligned}
\frac{\partial a}{\partial f_t}(w, \hat{f}_{n,t}, 0) &= \frac{\gamma}{\eta\sqrt{1-\gamma^2}}\phi\left(\frac{m^{-1}(w) - \mu - \gamma(\hat{f}_{n,t} - \mu)}{\eta\sqrt{1-\gamma^2}}\right), \\
\frac{\partial^2 a}{\partial f_t^2}(w, \hat{f}_{n,t}, 0) &= \frac{\gamma^2}{\eta^2(1-\gamma^2)}\left[\frac{m^{-1}(w) - \mu - \gamma(\hat{f}_{n,t} - \mu)}{\eta\sqrt{1-\gamma^2}}\right]\phi\left(\frac{m^{-1}(w) - \mu - \gamma(\hat{f}_{n,t} - \mu)}{\eta\sqrt{1-\gamma^2}}\right).
\end{aligned}$$

Then, from Proposition 7 and equations (6.13) and (6.22), the GA for filtering (6.26) follows.