MICROINFORMATION, NONLINEAR FILTERING AND GRANULARITY

Patrick GAGLIARDINI*, Christian GOURIEROUX† and Alain MONFORT‡

Preliminary Version: December 2009

*University of Lugano and Swiss Finance Institute.

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

‡CREST, Banque de France and Maastricht University.

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 functional 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 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.

Keywords: Kalman Filter, Nonlinear State Space, Granularity, Repeated Observations, Value-at-Risk, Basel 2.

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 distribution 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 distribution 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 probability distribution function (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 $p(\tilde{f}_t|\mathbf{Y}_t)$ of factor f_t given $\mathbf{Y}_t$, where the tilde notation is used to indicate the generic argument of a function. The

[1] This model is extended in Section 2 to allow for the effect of exogenous regressors and lagged observations in the measurement equation.

predictive density provides the conditional pdf of y_{t+1} given $\mathbf{Y}_t$, denoted $p(\tilde{y}_{t+1}|\mathbf{Y}_t)$. Then the joint pdf of the observations is deduced by multiplying the successive predictive densities, evaluated at the observed values $\tilde{y}_{t+1} = y_{t+1}$.

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

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

where:

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

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 set of function values $p(\tilde{f}_t|\mathbf{Y}_t)$, for any value of argument $\tilde{f}_t$. Hence, it is necessary to temporarily store the entire function at each recursion [2]. Three special cases are known, in which the nonlinear Kalman filter is simplified, because

[2]This is usually considered as an issue of numerical approximation. However, it may not be easy to

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 a Value-at-Risk (VaR) for a large homogeneous portfolio is discussed in Section 5. Section 6 concludes. Proofs are gathered in appendices. For simplicity, we focus on the most

impose the appropriate shape restrictions to ensure that the approximate function is indeed a legitimate pdf, and to control the associated approximation error.

common case of a single factor. Generalizing the results to multiple factors is theoretically straightforward, but notationally cumbersome.

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. 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.

State equation: *The conditional distribution of f_t given $\mathbf{F}_{t-1}$, $\mathbf{Y}_{t-1}$, X depends on f_{t-1}*

[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 $z_{i,t}$, z_t, say, in the measurement equations. However, as usual in state space models, the filtering and predictive distributions at horizon larger than 1 require the specification of the dynamics of the variables $z_{i,t}$ and z_t.

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 $h_{i,t}(y_{i,t}|f_t) = h(y_{i,t}|f_t)$.

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 Propositon 1 (see Appendix 1 for the proof). This result extends the approximate filtering distribution derived in Gagliardini and Gouriéroux (2009b) to a model with micro-dynamics and exogeneous 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. the filtering distribution. This distribution is Gaussian:*

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

At order $1/n$, the filtering distribution of f_t differs from a point mass at the CSML estimate $\hat{f}_{n,t}$. More precisely, the variance of the 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 adjustment involves the four summary statistics $\hat{f}_{n,t}$, $\hat{f}_{n,t-1}$, $I_{n,t}$, $K_{n,t}^{(3)}$, and no longer the lagged factor values. Conditionally on $\mathbf{Y}_t$ and X, f_t and $\mathbf{F}_{t-1}$ are independent at order $1/n$. The dynamics of the latent factor impacts the filtering distribution through the partial derivative of the log transition pdf $\frac{\partial \log g}{\partial f_t}(\hat{f}_{n,t}|\hat{f}_{n,t-1})$.

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

[4] See Bates (2009), p. 25, for approximations written on the same conditional distribution as ours. These

are typically unknown [5], while in our approch 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.

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:

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

where $\Psi(\tilde{y}_{t+1}|f_t, y_t, X) = p\left(\tilde{y}_{t+1}|\mathbf{Y}_t, \mathbf{F}_t, X\right)$ depends on the past through f_t, y_t only because of the assumptions on the state and measurement equations. Thus, the derivation of the

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.

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 distribution given $\mathbf{Y}_t$ and X in Proposition 1.

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

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

where $\tilde{y}_{i,t+1}$ and $\tilde{f}_{t+1}$ denote the arguments of the density functions, whereas f_t, y_t and X correspond to the variables in the information set. This density can be written as:

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

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 $1/n$ we have:*

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

The normalization factor $\sqrt{2\pi/n}$ ensures that the integral of $\Psi(\tilde{y}_{t+1}|f_t, y_t, X)$ 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.

Finally, the expression of Proposition 2 can be integrated w.r.t. the approximate Gaussian filtering distribution of f_t given in Proposition 1. We get the following result:

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

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

We get a closed form expression for the predictive distribution. 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:

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

Therefore, the adjustment at order $1/n$ in the filtering distribution (Proposition 1) involving the third-order derivative of the micro-density corresponds to a measure of conditional skewness through statistic $I_{n,t}^{-3/2} K_{n,t}^{(3)}$. 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 the 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, where the canonical factor 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, where the canonical factor f_t is related with the conditional probability p_t by $f_t = \log\left[p_t/(1-p_t)\right]$. The CSML estimator of the factor value is $\hat{f}_{n,t} = \log\left[\bar{y}_{n,t}/(1-\bar{y}_{n,t})\right]$, 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 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 some 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 Gaussian autoregressive factor:

$$g(f_t|f_{t-1};\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}\left(f_t - \mu - \rho f_{t-1}\right)^2\right], \tag{4.1}$$

where the macro-parameter $\theta = (\mu, \rho, \sigma^2)'$ is unknown. 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.

4.1 Approximate Log-likelihood Function

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

$$\mathcal{L}_{nT}(\theta) = \mathcal{L}_{nT}^{\mathrm{GA}}(\theta) + o_p(1/n), \tag{4.2}$$

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

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

and $p(y_t|\mathbf{Y}_{t-1}, X; \theta)$ is the approximate predictive distribution 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:

$$\hat{\theta}_{nT}^{\mathrm{GA}} = \arg\max_{\theta} \mathcal{L}_{nT}^{\mathrm{GA}}(\theta). \tag{4.3}$$

The granularity adjusted estimator differs from the unfeasible maximum likelihood estimator by a term negligible at order $1/n$ [see Gagliardini, Gouriéroux (2009a)]:

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

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

$$\mathcal{L}_{nT}^{\mathrm{GA}}(\theta) = \mathcal{L}_{0,nT} + \frac{1}{T}\sum_{t=1}^{T} \log p_1(y_t|\mathbf{Y}_{t-1}, X; \theta), \tag{4.4}$$

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

$$\begin{aligned}\log p_1(y_t|\mathbf{Y}_{t-1},X;\theta) &= \log g\left(\hat{f}_{n,t}|\hat{f}_{n,t-1};\theta\right)\\
&+\frac{1}{2n}I_{n,t}^{-1}\left(\frac{\partial^2 \log g\left(\hat{f}_{n,t}|\hat{f}_{n,t-1};\theta\right)}{\partial f_t^2}+\left(\frac{\partial \log g\left(\hat{f}_{n,t}|\hat{f}_{n,t-1};\theta\right)}{\partial f_t}\right)^2\right)\\
&+\frac{1}{2n}I_{n,t-1}^{-1}\left(\frac{\partial^2 \log g\left(\hat{f}_{n,t}|\hat{f}_{n,t-1};\theta\right)}{\partial f_{t-1}^2}+\left(\frac{\partial \log g\left(\hat{f}_{n,t}|\hat{f}_{n,t-1};\theta\right)}{\partial f_{t-1}}\right)^2\right)\\
&+\frac{1}{n}I_{n,t-1}^{-1}\frac{\partial \log g\left(\hat{f}_{n,t}|\hat{f}_{n,t-1};\theta\right)}{\partial f_{t-1}}\frac{\partial \log g\left(\hat{f}_{n,t-1}|\hat{f}_{n,t-2};\theta\right)}{\partial f_{t-1}}\\
&+\frac{1}{2n}I_{n,t}^{-2}K_{n,t}^{(3)}\frac{\partial \log g\left(\hat{f}_{n,t}|\hat{f}_{n,t-1};\theta\right)}{\partial f_t}+\frac{1}{2}I_{n,t-1}^{-2}K_{n,t-1}^{(3)}\frac{\partial \log g\left(\hat{f}_{n,t}|\hat{f}_{n,t-1};\theta\right)}{\partial f_{t-1}}.\end{aligned} \tag{4.5}$$

When the transition p.d.f. of the factor corresponds to the Gaussian autoregressive model (4.1), 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 (4.5) are polynomials in $\hat{f}_{n,t}-\mu-\rho\hat{f}_{n,t-1}$ and $\hat{f}_{n,t-1}-\mu-\rho\hat{f}_{n,t-2}$ 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 $\hat{f}_{n,t}-\mu-\rho\hat{f}_{n,t-1}$, $t=1,\cdots,T$, 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:*

$$\xi_{n,t} = \mu + \rho \xi_{n,t-1} + \sigma \varepsilon_t + \frac{1}{\sqrt{n}} I_{n,t}^{-1/2} u_t - \rho \frac{1}{\sqrt{n}} I_{n,t-1}^{-1/2} u_{t-1}, \quad t = 1, \cdots, T, \tag{4.6}$$

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 (4.6) 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 $\Omega_n = \sigma^2 Id_T + \frac{1}{n} B_n$, where B_n is the symmetric (T,T) matrix with elements equal to $I_{n,t}^{-1} + \rho^2 I_{n,t-1}^{-1}$ in position (t,t), $-\rho I_{n,t-1}^{-1}$ in positions $(t-1,t)$ and $(t,t-1)$, and zeros otherwise. At order $1/n$, we have:

$$\Omega_n^{-1} = \frac{1}{\sigma^2} Id_T - \frac{1}{n\sigma^4} B_n. \tag{4.7}$$

4.2 Approximate Linear Kalman Filter

We give below an equivalent statement of Propostion 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:*

$$f_t = \mu + \rho f_{t-1} + \sigma\varepsilon_t, \ \varepsilon_t \sim IIN(0,1), \tag{4.8}$$

and measurement equation:

$$\xi_{n,t} = f_t + \frac{1}{\sqrt{n}} I_{n,t}^{-1/2} u_t, \ u_t \sim IIN(0,1), \tag{4.9}$$

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

By replacing f_t in (4.8) by its expression derived from (4.9), we recover the recursive equation (4.6) in Proposition 5. Equivalently, (4.8)-(4.9) is the linear state space representation of the ARMA(1,1) process of Proposition 5. The granularity adjustment in the measurement equation (4.9) 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 less 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 (4.7).

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 assets. Its future value can be written as:

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

where the individual asset values $y_{i,t+1}$, $i = 1, \cdots, n$, are assumed to satisfy the assumptions of the nonlinear state space model in Section 2.1, with underlying factor value f_{t+1}. For expository purpose, we include neither exogeneous variables nor lagged observations in the measurement equations. The VaR at risk level α, with $\alpha \in (0, 1)$, 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. It is usual in this framework 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 $\mathrm{VaR}_{n,t}(\alpha)$, say, depends on the portfolio size and on the information 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 Value

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

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

where $m(f_{t+1}) = E[y_{i,t+1}|f_{t+1}]$, $\sigma^2(f_{t+1}) = V[y_{i,t+1}|f_{t+1}]$ and Z is a standard Gaussian variable independent of $\mathbf{F}_{t+1}, \mathbf{Y}_t$. The relation (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 (since $W_{n,t+1}/n$ is an unbiased estimator of $m(f_{t+1})$ conditional on f_{t+1}). The approximate cdf of $W_{n,t+1}/n$ given f_{t+1} is such that [7] $P\left[W_{n,t+1}/n \leq w|f_{t+1}\right] = \Phi\left(\frac{w - m(f_{t+1})}{\sigma(f_{t+1})/\sqrt{n}}\right) + o(1/n)$.

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

$$\begin{aligned} P\left[W_{n,t+1}/n \leq w|\mathbf{F}_t, \mathbf{Y}_t, Z\right] &= \int \mathbb{I}_{m(f_{t+1}) + \frac{\sigma(f_{t+1})}{\sqrt{n}} Z \leq w} g(f_{t+1}|f_t) df_{t+1} + o(1/n) \\ &= a(w, f_t, Z/\sqrt{n}) + o(1/n), \text{ say.} \end{aligned} \tag{5.3}$$

Under mild regularity conditions, function $a(w, f, \varepsilon)$ is continuously differentiable w.r.t. the

[7] The approximation error $o(1/n)$ is derived by writing the cdf of the conditional distribution of $W_{n,t+1}/n$ given f_{t+1} as an integral of the associated complex Laplace transform by means of the Fourier Inversion formula [Duffie, Pan, Singleton (2000)]. This Laplace transform corresponds to a Gaussian distribution at order $o(1/n)$ by the CLT.

arguments f and ε at $\varepsilon = 0$ (see below).

(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:

$$
\begin{aligned}
F_{n,t}(w) &= E\left[a(w, f_t, Z/\sqrt{n})|\mathbf{Y}_t\right] + o(1/n) \\
&= E\left[a\left(w, \hat{f}_{n,t} + \frac{1}{n}\mu_{n,t} + \frac{1}{\sqrt{n}}I_{n,t}^{-1/2}Z^*, \frac{1}{\sqrt{n}}Z\right)\right] + o(1/n), \qquad (5.4)
\end{aligned}
$$

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

Then, we can expand equation (5.4) at order $1/n$. Since $E[Z] = E[Z^*]$ =0, $E[ZZ^*] = 0$, $E[Z^2] = E[(Z^*)^2] = 1$, we get:

$$
\begin{aligned}
F_{n,t}(w) &= a(w, \hat{f}_{n,t}, 0) + \frac{1}{n}\frac{\partial a}{\partial f}(w, \hat{f}_{n,t}, 0)\mu_{n,t} \\
&\quad + \frac{1}{2n}\left[I_{n,t}^{-1}\frac{\partial^2 a}{\partial f^2}(w, \hat{f}_{n,t}, 0) + \frac{\partial^2 a}{\partial \varepsilon^2}(w, \hat{f}_{n,t}, 0)\right] + o(1/n). \qquad (5.5)
\end{aligned}
$$

In the above expression we distinguish three components:

$a(w, \hat{f}_{n,t}, 0) =: F_{\infty,t}(w)$ 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 $\hat{f}_{n,t}$;

a first GA $\frac{1}{2n}\frac{\partial^2 a}{\partial \varepsilon^2}(w, \hat{f}_{n,t}, 0)$ 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. Moreover, 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].

5.3 Granularity Adjustment of the standardized VaR

Finally, the GA of the VaR is directly deduced from (5.5) by applying the Bahadur's expansion. Let us denote $Q_{n,t}$ (resp. $Q_{\infty,t}$) the quantile function corresponding to $F_{n,t}$ (resp. $F_{\infty,t}$), and assume that the cross-sectional asymptotic density $f_{\infty,t}(w) = dF_{\infty,t}(w)/dw$ exists and is strictly positive. We have [Bahadur (1966)]:

$$Q_{n,t}(\alpha) - Q_{\infty,t}(\alpha) = -\frac{F_{n,t}\left[Q_{\infty,t}(\alpha)\right] - \alpha}{f_{\infty,t}\left[Q_{\infty,t}(\alpha)\right]} + o(1/n). \tag{5.6}$$

The GA for the quantile and for the standardized VaR are obtained by replacing $F_{n,t}$, $F_{\infty,t}$, ... by their expressions using (5.5). 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 invertibility conditions, the derivatives of function $a(w, f, \varepsilon)$ w.r.t. to ε at 0 can also be expressed in terms of the conditional distributions defining the measurement and state equations. First, let us assume that function $m(.)$ is one-to-one. Then, up to a (nonlinear) transformation of the factor we can assume that $m(f) = f$, that is, the factor f_t is identified with the conditional mean of the individual observations. Moreover, let the factor f_t admit values in set $\mathcal{F} \subset \mathbb{R}$, and define the function:

$$\Psi(f, \varepsilon) = f + \varepsilon\sigma(f),$$

for $f \in \mathcal{F}$ and $\varepsilon \in \mathbb{R}$.

Assumption 1: *For any ε in a neighbourhood of 0, the inverse of function $f \to \Psi(f, \varepsilon)$, denoted by $\Psi^{-1}(\cdot, \varepsilon)$, is well defined on $\mathcal{F}$ and is such that*

$$\Psi(f, \varepsilon) \leq w \ \Leftrightarrow f \leq \Psi^{-1}(w, \varepsilon),$$

for any $w, f \in \mathcal{F}$.

Intuitively, Assumption 1 is satisfied when the effect of the volatility function $\sigma(f)$ is bounded. We illustrate the (non) validity of Assumption 1 in some examples in Section 5.4. Under Assumption 1 the function a becomes:

$$a(w, f_t, \varepsilon) = \int_{-\infty}^{\Psi^{-1}(w,\varepsilon)} g(f_{t+1}|f_t)df_{t+1}. \tag{5.7}$$

Then, function a is differentiable w.r.t. ε at $\varepsilon = 0$ and we get the following Proposition (see Appendix 6 for the proof):

PROPOSITION 7: *If $m(f) = f$ and Assumption 1 is satisfied, the GA for the finite portfolio size is:*

$$-\frac{1}{2n}\sigma^2(Q_{\infty,t}(\alpha))\left(\frac{\partial \log g}{\partial f_{t+1}}\left(Q_{\infty,t}(\alpha)|\hat{f}_{n,t}\right)+\frac{d\log\sigma^2}{df}\left(Q_{\infty,t}(\alpha)\right)\right),$$

and the GA for filtering the current factor value is:

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

For a static factor model, the GA for filtering the current factor value is equal to zero, and the GA for finite portfolio size becomes:

$$-\frac{1}{2n}\sigma^2(Q_{\infty,t}(\alpha))\frac{\partial \log\left(g\cdot\sigma^2\right)}{\partial f}\left(Q_{\infty,t}(\alpha)\right).$$

This formula corresponds to the GA derived in Martin, Wilde (2001), Gordy (2004) following the local analysis of VaR in Gouriéroux, Laurent, Scaillet (2000). 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) and Gaussian transformed factor.

i) Linear Gaussian state space model

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

$$y_{i,t} = \mu + \eta\sqrt{\rho}F_t + \eta\sqrt{1-\rho}u_{i,t}, \ \ i = 1, ..., n, \tag{5.8}$$

and state equation:

$$F_t = \gamma F_{t-1} + \sqrt{1-\gamma^2}v_t, \tag{5.9}$$

where $(u_{i,t})$, $i = 1, ..., n$, and (v_t) are independent $IIN(0, 1)$ processes. The factor F_t is normalized to have $N(0, 1)$ stationary distribution, with autoregressive parameter γ such that $|\gamma| < 1$. The unconditional distribution of the variables $y_{i,t}$ is $N(\mu, \eta^2)$, and the parameter $\rho > 0$ is the unconditional correlation between any pair $(y_{i,t}, y_{j,t})$, for $i \neq j$. The conditional distribution of $y_{i,t}$ given F_t is Gaussian $N(f_t, \sigma^2)$, with conditional mean corresponding to the transformed factor $f_t = \mu + \eta\sqrt{\rho}F_t$, and idiosyncratic conditional variance $\sigma^2 = (1-\rho)\eta^2$ independent of the factor value. By using that the distribution of f_{t+1} conditional on f_t is $N\left(\mu + \gamma(f_t - \mu), \rho\eta^2(1-\gamma^2)\right)$, we deduce:

$$a(w, f, 0) = \Phi\left(\frac{w - \mu - \gamma(f-\mu)}{\eta\sqrt{\rho}\sqrt{1-\gamma^2}}\right).$$

By inversion w.r.t. w, we get the quantile per individual asset computed on a portfolio of infinite size:

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

Let us now derive the GA. We have $m(f) = f$ and $\sigma(f) = \sigma$ for $f \in \mathcal{F} = \mathbb{R}$, and Assumption 1 is satisfied with $\Psi^{-1}(w, \varepsilon) = w - \varepsilon\sigma$. Then, from Proposition 7 the GA the finite portfolio size is:

$$\frac{1}{2n}\frac{\sigma^2}{\eta\sqrt{\rho}\sqrt{1-\gamma^2}}\Phi^{-1}(\alpha),$$

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

$$\frac{\eta\gamma}{n\sqrt{\rho}\sqrt{1-\gamma^2}}\left[\frac{1}{2}\gamma\Phi^{-1}(\alpha) - \hat{v}_{n,t}\right],$$

where $\hat{v}_{n,t} = \dfrac{\hat{f}_{n,t} - \mu - \gamma(\hat{f}_{n,t-1} - \mu)}{\eta\sqrt{\rho}\sqrt{1-\gamma^2}}$ is the filtered 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 unit 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 (5.8)-(5.9). 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 $\mathcal{B}(1, f_t)$ conditional on the factor, with conditional default probability:

$$f_t = P\left[y_{i,t} = 1|F_t\right] = P\left[\mu + \eta\sqrt{\rho}F_t + \eta\sqrt{1-\rho}u_{i,t} < 0|F_t\right] = \Phi\left(-\frac{\mu + \eta\sqrt{\rho}F_t}{\eta\sqrt{1-\rho}}\right).$$

The state equation is such that the transition pdf of transformed factor f_t is deduced from the Gaussian transition pdf of F_t in (5.9).

Let us first compute the function $a(w, f, 0)$. Since $F_{t+1} = -\frac{\mu + \eta\sqrt{1-\rho}\Phi^{-1}(f_{t+1})}{\eta\sqrt{\rho}}$, we have:

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

By inverting this function w.r.t. w, we get the predictive individual quantile at level α computed on a portfolio of infinite size:

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

Let us now derive the GA of the quantile. We have $m(f) = f$ and $\sigma(f) = \sqrt{f(1-f)}$ for $f \in \mathcal{F} = [0, 1]$. For any $\varepsilon \in \mathbb{R}$ and $w \in [0, 1]$, the equation $f + \varepsilon\sqrt{f(1-f)} = w$, for $f \in [0, 1]$, admits the unique solution

$$f = \Psi^{-1}(w, \varepsilon) = \frac{2w + \varepsilon^2 - \varepsilon\sqrt{4w(1-w) + \varepsilon^2}}{2(1+\varepsilon^2)}.$$

Assumption 1 is satisfied since function $w \to \Psi^{-1}(w, \varepsilon)$ is monotone on $[0, 1]$. From Proposition 7, the GA for finite portfolio size is:

$$\frac{1}{2n}\left\{\left[\frac{\sqrt{1-\rho}}{\sqrt{\rho}\sqrt{1-\gamma^2}}\Phi^{-1}(\alpha) - \Phi^{-1}(Q_{\infty,t}(\alpha))\right]\frac{Q_{\infty,t}(\alpha)[1-Q_{\infty,t}(\alpha)]}{\phi\left[\Phi^{-1}(Q_{\infty,t}(\alpha))\right]} - 1\right\},$$

and the GA for filtering the current factor value is:

$$\frac{\gamma}{n}\phi\left[\Phi^{-1}(Q_{\infty,t}(\alpha))\right]\left(\mu_{n,t}\Phi^{-1}(\hat{f}_{n,t}) + \frac{1}{2}\frac{I_{n,t}^{-1}}{\phi\left[\Phi^{-1}(\hat{f}_{n,t})\right]} + \frac{\gamma}{2}\frac{\sqrt{1-\rho}}{\sqrt{\rho}\sqrt{1-\gamma^2}}I_{n,t}^{-1}\Phi^{-1}(\alpha)\left[\Phi^{-1}(\hat{f}_{n,t})\right]^2\right).$$

iii) Gaussian model with volatility factor

Let us assume that the observations $y_{i,t}$, $i = 1, \cdots, n$, are $IIN(0, f_t^2)$ distributed conditional on the volatility factor $f_t > 0$. For this model the conditional mean function $m(f) = 0$ is independent of the factor value. Moreover, the invertibility condition corresponding to Assumption 1 is not satisfied, since for instance for any $w < 0$ and $\varepsilon > 0$ we have $\Psi(f, \varepsilon) = \varepsilon f \leq w$ for no $f \geq 0$. From (5.3) we have for $w < 0$:

$$a(w, f_t, \varepsilon) = \begin{cases} 0, & \text{if } \varepsilon \geq 0 \\ S\left(\frac{w}{\varepsilon}|f_t\right), & \text{if } \varepsilon < 0 \end{cases},$$

where $S(f_{t+1}|f_t)$ denotes the transition survivor function of the factor. The right second-order derivative of function $a(w, f, \varepsilon)$ at $\varepsilon = 0$ is:

$$\begin{aligned}\lim_{\varepsilon\to 0^-}\frac{\partial^2 a}{\partial\varepsilon^2}(w, f_t, \varepsilon) &= \lim_{\varepsilon\to 0^-}\left(-\frac{\partial g(w/\varepsilon|f_t)}{\partial f_{t+1}}(w/\varepsilon^2)^2 - 2g(w/\varepsilon|f_t)w/\varepsilon^3\right) \\ &= \frac{1}{w^2}\lim_{z\to\infty}\left(-\frac{\partial g(z|f_t)}{\partial f_{t+1}}z^4 - 2g(z|f_t)z^3\right).\end{aligned}$$

Thus, if the tails of the transition distribution of the factor are sufficiently thin such that $\lim_{z\to\infty}\left(-\frac{\partial g(z|f_t)}{\partial f_{t+1}}z^4 - 2g(z|f_t)z^3\right) = 0$, then function $a(w, f_t, \varepsilon)$ is twice-differentiable at $\varepsilon = 0$, with zero second-order derivative. Otherwise, if the tails are not thin, the function $a(w, f_t, \varepsilon)$ is not twice-differentiable at $\varepsilon = 0$.

As an example, let us assume that the transition density of the factor is exponential, with survivor function $S(z|f_t) = \exp(-\lambda(f_t)z)$, $\lambda(f_t) > 0$ [8]. Then, from (5.3) the conditional cdf of $W_{n,t+1}/n$ given f_t at order $1/n$ is:

$$E\left[a(w, f_t, Z/\sqrt{n})\right] = E\left[\exp\left(-\frac{w\lambda(f_t)}{\sqrt{n}}Z\right)\mathbb{I}_{Z<0}\right] = \Phi\left(\frac{\lambda(f_t)w}{\sqrt{n}}\right)\exp\left(\frac{1}{2n}\lambda(f_t)^2w^2\right).$$

This formula can be expanded at order $1/n$ to get:

$$E\left[a(w, f_t, Z/\sqrt{n})\right] = \frac{1}{2} + \frac{w\lambda(f_t)}{\sqrt{2\pi n}} + \frac{\lambda(f_t)^2w^2}{4n} + o(1/n).$$

By integrating out the factor f_t using the approximate filtering distribution in Proposition 1, we get:

$$F_{n,t}(w) = \frac{1}{2} + \frac{w\lambda(\hat{f}_{n,t})}{\sqrt{2\pi n}} + \frac{\lambda(\hat{f}_{n,t})^2w^2}{4n} + o(1/n).$$

Equivalently:

$$F_{n,t}(w) = \Phi\left(\frac{\lambda(\hat{f}_{n,t})w}{\sqrt{n}}\right)\exp\left(\frac{1}{2n}\lambda(\hat{f}_{n,t})^2w^2\right) + o(1/n).$$

[8] In this example, the derivatives at any order of function $a(w, f_t, \varepsilon)$ w.r.t. ε in $\varepsilon = 0$ are zero. Since the function is not equal to zero for $\varepsilon < 0$, the convergence radius of the Taylor series around $\varepsilon = 0$ is zero. Thus, we cannot apply the argument in (5.5) based on a Taylor expansion.

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

[5] Duffie, D., Pan, J., and K., Singleton (2000): ”Transform Analysis and Asset Pricing for Affine Jump Diffusions”, Econometrica, 68, 1343-1376.

[6] Gagliardini, P., and C., Gouriéroux (2009a): “Efficiency in Large Dynamic Panel Models with Common Factor”, CREST DP.

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

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

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

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

[11] Gouriéroux, C., and A., Monfort (2009): "Granularity in Qualitative Factor Model", forthcoming in the Journal of Credit Risk.

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

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

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

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

[16] Kitagawa, G. (1987): "Non-Gaussian State Space Modelling of Nonstationary Time Series", JASA, 82, 1032-1063.

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

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

[19] Martin, Wilde (2002): "Unsystematic Credit Risk", Risk, 15, 123-128.

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

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

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

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

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

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:

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

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 (2009b), we get:

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

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 $\tilde{f}_{t+1} = \tilde{f}_{n,t+1}$. We have:

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

Let us introduce the change of variable:

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

Then, we get:

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

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

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

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:

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

where $\Psi\left(\tilde{y}_{t+1}|f_t, y_t, X\right)$ 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:

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

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:

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

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

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

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:

$$E\left[\exp\left(\frac{1}{\sqrt{n}}\left[I_{n,t}^{-1/2}\frac{\partial\log g\left(\tilde{f}_{n,t+1}|\hat{f}_{n,t}\right)}{\partial f_t}+I_{n,t}^{1/2}\xi_{n,t}\right]Z+\frac{I_{n,t}^{-1}}{2n}\frac{\partial^2\log g\left(\tilde{f}_{n,t+1}|\hat{f}_{n,t}\right)}{\partial f_t^2}Z^2\right)\right]$$

$$= \exp\left\{\frac{1}{2n}\left[I_{n,t}^{-1}\frac{\partial^2\log g\left(\tilde{f}_{n,t+1}|\hat{f}_{n,t}\right)}{\partial f_t^2}+\left(I_{n,t}^{-1/2}\frac{\partial\log g\left(\tilde{f}_{n,t+1}|\hat{f}_{n,t}\right)}{\partial f_t}+I_{n,t}^{1/2}\xi_{n,t}\right)^2\right]+o(1/n)\right\}.$$

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:

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

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

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

and:

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

(ii) The second-order derivative is:

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

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

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

(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 $E\left[\frac{1}{h_{i,t}\left(y_{i,t}|f_t\right)}\frac{\partial^3 h_{i,t}\left(y_{i,t}|f_t\right)}{\partial f_t^3}|f_t\right] = 0$, $E\left[\frac{\partial \log h_{i,t}\left(y_{i,t}|f_t\right)}{\partial f_t}|f_t\right] = 0$ 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:

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

(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:

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

By substituting:

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

we get:

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

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

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

Appendix 5: Proof of Proposition 5

Let us first rewrite the RHS of (4.5). By using $\log g(\hat{f}_{n,t}|\hat{f}_{n,t-1};\theta) = -\frac{1}{2}\log(2\pi\sigma^2) - \frac{\hat{\eta}_{n,t}(\theta)^2}{2\sigma^2}$, $\frac{\partial \log g(\hat{f}_{n,t}|\hat{f}_{n,t-1};\theta)}{\partial f_t} = -\frac{\hat{\eta}_{n,t}(\theta)}{\sigma^2}$, $\frac{\partial^2 \log g(\hat{f}_{n,t}|\hat{f}_{n,t-1};\theta)}{\partial f_t^2} = -\frac{1}{\sigma^2}$, $\frac{\partial \log g(\hat{f}_{n,t}|\hat{f}_{n,t-1};\theta)}{\partial f_{t-1}} = \frac{\rho\hat{\eta}_{n,t}(\theta)}{\sigma^2}$ and $\frac{\partial^2 \log g(\hat{f}_{n,t}|\hat{f}_{n,t-1};\theta)}{\partial f_{t-1}^2} = -\frac{\rho^2}{\sigma^2}$, where $\hat{\eta}_{n,t}(\theta) = \hat{f}_{n,t} - \mu - \rho\hat{f}_{n,t-1}$, we get:

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

Thus from (4.4), the GA log-likelihood function can be written as:

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

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

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

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

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:

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

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 $\sigma\varepsilon_t + \frac{1}{\sqrt{n}}I_{n,t}^{-1/2}u_t - \rho\frac{1}{\sqrt{n}}I_{n,t-1}^{-1/2}u_{t-1}$, where (ε_t) and (u_t) are independent Gaussian white noise processes, the conclusion follows.

Appendix 6: Proof of Proposition 7

From (5.7), the second-order derivative w.r.t. ε at $\varepsilon = 0$ is given by:

$$\frac{\partial^2 a}{\partial \varepsilon^2}(w, f_t, 0) = g\left(w|f_t\right) \left[\frac{\partial \log g\left(w|f_t\right)}{\partial f_{t+1}} \left(\frac{\partial \Psi^{-1}(w,0)}{\partial \varepsilon} \right)^2 + \frac{\partial^2 \Psi^{-1}(w,0)}{\partial \varepsilon^2} \right],$$

where we have used that $\Psi^{-1}(w,0) = w$. To compute the partial derivatives of $\Psi^{-1}(w,\varepsilon)$ w.r.t. ε at 0, we differentiate twice w.r.t. ε the identity $\Psi\left(\Psi^{-1}(w,\varepsilon),\varepsilon\right) = w$, and evaluate the result at $\varepsilon = 0$. We get

$$\frac{\partial \Psi^{-1}(w,0)}{\partial \varepsilon} = -\frac{\partial \Psi(w,0)}{\partial \varepsilon} \Big/ \frac{\partial \Psi(w,0)}{\partial f} = -\sigma(w),$$

and:

$$\begin{aligned} \frac{\partial^2 \Psi^{-1}(w,0)}{\partial \varepsilon^2} &= -\left(\frac{\partial \Psi(w,0)}{\partial f} \right)^{-1} \left[\frac{\partial^2 \Psi(w,0)}{\partial f^2} \left(\frac{\partial \Psi^{-1}(w,0)}{\partial \varepsilon} \right)^2 + 2 \frac{\partial^2 \Psi(w,0)}{\partial f \partial \varepsilon} \frac{\partial \Psi^{-1}(w,0)}{\partial \varepsilon} + \frac{\partial^2 \Psi(w,0)}{\partial \varepsilon^2} \right] \\ &= \frac{d\sigma^2(w)}{df}. \end{aligned}$$

We deduce:

$$\frac{\partial^2 a}{\partial \varepsilon^2}(w, f_t, 0) = g\left(w|f_t\right) \sigma^2(w) \left(\frac{\partial \log g\left(w|f_t\right)}{\partial f_{t+1}} + \frac{d \log \sigma^2(w)}{df} \right).$$

By using equations (5.5) and (5.6), the conclusion follows.

Table 1: Canonical parameters and summary statistics in the main 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 $N(0,\sigma^2)$	$f = 1/\sigma^2$	$\hat{f}_{n,t} = 1/\hat{\sigma}^2_{n,t}$	$c(f) = \log f/2$	$a(y) = -y^2/2$

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.