Essays on n-Nonlinear Panel Time Series Models

Mesters, G.

2015

Chapter 1

Introduction

This thesis develops methods for conducting inference on nonlinear panel data models in the presence of latent stochastic variables. We focus on developing likelihood-based methods that enable the efficient estimation of the deterministic model parameters and allow for the extraction of conditional estimates for the latent stochastic variables. The new estimation methods are based on either simulation methods, such as importance sampling, or on iterative optimization methods, such as the expectation-maximization algorithm. The methodology is illustrated using panel data from different fields of research including macroeconomics, microeconomics, finance and sociology.

We focus on four specific types of deviations from the standard linear Gaussian panel data model. First, we consider the case where the observations follow non-Gaussian distributions. This situation is commonly encountered in empirical applications in micro-econometrics, where we observe binary and count panel data; see Cameron & Trivedi (2005, Chapter 23). Also, in risk-management and finance we find data panels that are more appropriately modeled by skewed and heavy tailed distributions; see McNeil, Frey & Embrechts (2005).

Second, we consider observations for which the mean depends on a nonlinear combination of latent stochastic variables. In particular, we consider the case where the mean of the observations is determined by the inner product of two stochastic vectors, where one of the vectors varies stochastically over time. This dynamic factor structure allows for dimension reduction in panel data and provides an attractive specification for forecasting in situations where there are many possible predictors; see Bai & Ng

(2008) and Stock & Watson (2011). Applications for dynamic factor models exist in macroeconomics, labor economics and finance.

Third, we consider situations where the distribution of the observations is specified by a mean and variance that both depend on latent stochastic variables. Applications for these models arise in finance and macroeconomics, where both the mean and variance are dependent on the business cycle and financial conditions. Further, since observations from financial time series are more appropriately modeled by conditionally heavy-tailed densities, we combine the stochastic specifications for the mean and variance with Student's t distributions for the observations. In this manner we are able to model many features of financial data; see also Chib, Nardari & Shephard (2006).

Fourth and finally, we consider the case where the latent stochastic variables follow a long-memory process. When time series display long-memory they cannot be written in a finite state space form. While this is not an entirely nonlinear model specification, the latent long-memory variables do prevent the use of standard filtering methods for the estimation of the model parameters; see Chan & Palma (1998). Also, in order to generalize the model specification we consider non-Gaussian densities for the observations that are defined conditional on latent long-memory processes; see Brockwell (2007).

When we deviate from the standard linear Gaussian panel data model in any of the ways described above, the likelihood can no longer be expressed in closed form. For the evaluation of the likelihood we present new methods that are based on either the importance sampling technique, or the expectation-maximization algorithm. While these methods have been well developed for models with low cross-section dimensions, panel data structures present new challenges for modeling nonlinear features in the presence of latent stochastic variables. In particular, high dimensions for both cross-section and time require the development of new methods for conducting inference on the model parameters.

From an empirical perspective the research aims of this thesis can be formulated as follows. First, we are interested in flexible panel data models that allow for the modeling of a variety of non-Gaussian data structures that include binary, count and heavy-tailed data. In these situations we want to allow for heterogeneity in the model by including individual-specific and time-varying effects. Controlling for unobserved heterogeneity is important in panel data setting as often large differences exist between the individuals and the time periods. Second, we are interested in forecasting from nonlinear models. Forecasting using panel data has the advantage that co-movements between different time series can be exploited. Third, we aim to document the contemporaneous and long term interaction between the non-Gaussian time series by studying the interaction between the latent stochastic variables that capture the moments of the non-Gaussian densities.

1.1 Linear Panel Data Models

Consider the situation where we observe variables $y_{i,t}$ for a total of N individuals and T time periods. For each of the N individuals the observations are indexed by i, for $i = 1, \ldots, N$, and the time periods are indexed by t, for $t = 1, \ldots, T$. The term "individual" can refer to countries, firms, groups, persons or other separately definable entities. We are interested in situations where there are many individuals and many time periods. The standard linear Gaussian panel data model for the observations $y_{i,t}$ is given by

$$y_{i,t} = \boldsymbol{x}'_{i,t}\boldsymbol{\beta} + \mu_i + \xi_t + \epsilon_{i,t}, \qquad \epsilon_{i,t} \sim NID(0, \sigma^2), \tag{1.1}$$

where $\boldsymbol{x}_{i,t}$ is the $k \times 1$ vector of observable explanatory variables, $\boldsymbol{\beta}$ is the $k \times 1$ vector of parameters, μ_i is the individual-specific effect, ξ_t is the time-varying effect and $\epsilon_{i,t}$ is the disturbance term, which is normally and independently distributed with mean zero and variance σ^2. This is the basic panel data model that is discussed extensively in Baltagi (2005), where it is referred to as the two-way error panel data model.

Typically, interest is in conducting inference on the parameter vector $\boldsymbol{\beta}$, which measures the effect of the explanatory variables $\boldsymbol{x}_{i,t}$ on the observations $y_{i,t}$. However, in this thesis we are also, if not predominantly, interested in the individual-specific and time-varying effects; μ_i and ξ_t. We show that creative use of these variables allows for more flexible model specifications.

In the basic model (1.1) the individual-specific effects μ_i serve to capture the time-invariant differences between the time series. In other words, they capture the individual means of the time series. The time-varying effects ξ_t capture the time-varying differences that are common for all individuals. These can be viewed as the means of the time periods. When the vector $\boldsymbol{x}_{i,t}$ includes a constant the individual-specific and time-varying effects are to be seen as deviations from the common mean. The unobserved parameters μ_i and ξ_t can be considered either deterministic or stochastic.

In this thesis we consider the set-up where both the individual-specific and the time-varying effects are treated as latent stochastic variables. Two main reasons for this exist. First, the random variable set-up is more parsimonious since it avoids the estimation of N individual-specific and T time-varying effects. Instead filtering and smoothing methods are used to evaluate the conditional distributions of the individual-specific and time-varying effects. Second and most important, the random variable setup is more suitable for nonlinear panel data models. This follows as for nonlinear panel data models it is not always possible to find sufficient statistics that eliminate the fixed individual-specific effects. When this is not possible fixed-effects approaches suffer from the incidental parameter problem; see Neyman & Scott (1948) and Lancaster (2000). The random variable approach also has drawbacks. First, assumptions need to be made for the distributions of μ_i and ξ_t. Second, further assumptions need to be made for the relationship between the random effects and the explanatory variables; see Chamberlain (1980).

In the basic random effects panel data model the individual-specific and time-varying effects are modeled by Gaussian distributions. In particular, μ_i and ξ_t are given by

$$\mu_i \sim NID(0, \sigma_\mu^2), \qquad \xi_t \sim NID(0, \sigma_\xi^2), \tag{1.2}$$

where the distributions of μ_i, ξ_t and $\epsilon_{i,t}$ are typically considered independent. The specification for the random individual-specific effects causes the observations to be correlated among all time periods, whereas the specification for the random time-varying effects causes the observations to be correlated among all individuals. The parameters for the random effects panel data model (1.1) are collected in the parameter vector $\boldsymbol{\psi}$, which typically contains the parameters $\boldsymbol{\beta}$ and the parameters that pertain to the distributions of the random variables.

The log likelihood for model (1.1) is defined as $\ell(\boldsymbol{\psi}; \boldsymbol{y}) = \log p(\boldsymbol{y}; \boldsymbol{\psi})$, where $\boldsymbol{y} = \{y_{i,t}\}_{i=1,\dots,N,\ t=1,\dots,T}$ and $p(\boldsymbol{y}; \boldsymbol{\psi})$ denotes the joint density of the observations. Despite the complicated correlation structure for the observations, the likelihood for model (1.1) is available in closed form. In particular, the variance matrix for the stacked vector of observations $\boldsymbol{y} = (y_{1,1}, \dots, y_{1,T}, y_{2,1}, \dots, y_{2,T}, \dots, y_{N,1}, \dots, y_{N,T})'$ is given by

$$\boldsymbol{\Sigma}_y = \sigma_\mu^2(\boldsymbol{I}_N \otimes \boldsymbol{J}_T) + \sigma_\eta^2(\boldsymbol{J}_N \otimes \boldsymbol{I}_T) + \sigma_\epsilon^2(\boldsymbol{I}_N \otimes \boldsymbol{I}_T) \tag{1.3}$$

where $\boldsymbol{J}_k$ denotes the $k \times k$ matrix of ones and $\boldsymbol{I}_l$ denotes the $l \times l$ unit matrix. Given the Gaussian assumption for the disturbances the log likelihood can be expressed by

$$\ell(\boldsymbol{\psi}; \boldsymbol{y}) = \omega - \frac{1}{2}\log|\boldsymbol{\Sigma}_y| - \frac{1}{2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{\Sigma}_y^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}), \tag{1.4}$$

where ω is a constant and $\boldsymbol{X}$ is the $NT \times k$ matrix of stacked explanatory variables. Expressions for $\boldsymbol{\Sigma}_y^{-1}$ are given in Baltagi (2005), where also several methods for optimizing the log likelihood with respect to the parameters are discussed.

In many applications more flexibility can be introduced in the model by allowing for multiple individual-specific and time-varying effects. For example, in random coefficient panel data models all the deviations from the common regression parameters $\boldsymbol{\beta}$ are considered stochastic; see the discussion in Hsiao & Pesaran (2008). A generalization of model (1.1) that allows for multiple random effects is given by

$$y_{i,t} = \boldsymbol{x}_{i,t}'\boldsymbol{\beta} + \boldsymbol{c}_{i,t}'\boldsymbol{\mu}_i + \boldsymbol{d}_{i,t}'\boldsymbol{\xi}_t + \epsilon_{i,t}, \qquad \epsilon_{i,t} \sim NID(0, \sigma_\epsilon^2), \tag{1.5}$$

where $\boldsymbol{\mu}_i$ is a $q \times 1$ vector of individual-specific effects, which is weighted for individual i in time period t by the $q \times 1$ fixed vector $\boldsymbol{c}_{i,t}$ and $\boldsymbol{\xi}_t$ is a $r \times 1$ vector of time-varying effects, which is weighted for individual i in time period t by the $r \times 1$ fixed vector $\boldsymbol{d}_{i,t}$. The vectors $\boldsymbol{c}_{i,t}$ and $\boldsymbol{d}_{i,t}$ may include observable explanatory variables, known weights or unknown fixed parameters.

Correspondingly, the distributions of the random effects are adjusted as follows

$$\boldsymbol{\mu}_i \sim NID(\boldsymbol{0}, \boldsymbol{\Sigma}_\mu), \qquad \boldsymbol{\xi}_t \sim NID(\boldsymbol{0}, \boldsymbol{\Sigma}_\xi), \tag{1.6}$$

where $\boldsymbol{\Sigma}_\mu$ and $\boldsymbol{\Sigma}_\xi$ are the $q \times q$ and $r \times r$ variance matrices. Under these assumptions the likelihood remains available in closed form. However, the identification of the model parameters and the appropriate method for evaluating the likelihood does depend on the choice for the vectors $\boldsymbol{c}_{i,t}$ and $\boldsymbol{d}_{i,t}$.

In many applications it is likely that the time-varying effects are persistent over time. In other words, the assumption that the random time-varying effects are independent over time can be restrictive. Further, more advanced specifications for the time-varying effects may improve the out-of-sample predictive ability of the panel data model. Without compromising the closed form of the likelihood in (1.4) we may extend the specification for the time-varying effects in (1.6) by considering the vector autoregressive model given by

$$\boldsymbol{\xi}_{t+1} = \boldsymbol{\Phi}\boldsymbol{\xi}_t + \boldsymbol{\zeta}_t, \qquad \boldsymbol{\zeta}_t \sim NID(0, \boldsymbol{\Sigma}_\zeta), \qquad t = 1, \ldots, T, \tag{1.7}$$

where $\boldsymbol{\Phi}$ is the $r \times r$ autoregressive coefficient matrix and $\boldsymbol{\zeta}_t$ is the $r \times 1$ vector of innovations. For expositional purposes we assume that the vector autoregressive process in (1.7) has one lag and that it is stationary with mean zero and $\text{Var}(\boldsymbol{\xi}_t) = \boldsymbol{I}_r$. The vector autoregressive model (1.7) allows for the subsequent time-varying effects to depend on each other via the autoregressive matrix $\boldsymbol{\Phi}$.

When $\boldsymbol{\mu}_i = \mathbf{0}$ for all $i = 1, \ldots, N$, the model (1.5) with $\boldsymbol{\xi}_t$ as in (1.7) reduces to a linear Gaussian state space model; see Durbin & Koopman (2012, Part 1). To make this clear we rewrite the model as follows

$$\begin{aligned} \boldsymbol{y}_t &= \boldsymbol{X}_t\boldsymbol{\beta} + \boldsymbol{Z}_t\boldsymbol{\alpha}_t + \boldsymbol{\epsilon}_t, & \boldsymbol{\epsilon}_t &\sim NID(\mathbf{0}, \sigma^2\boldsymbol{I}_N), \\ \boldsymbol{\alpha}_{t+1} &= \boldsymbol{H}\boldsymbol{\alpha}_t + \boldsymbol{\eta}_t, & \boldsymbol{\eta}_t &\sim NID(\mathbf{0}, \boldsymbol{\Sigma}_\eta), \end{aligned} \tag{1.8}$$

where $\boldsymbol{y}_t = (y_{1,t}, \ldots, y_{N,t})'$ and $\boldsymbol{Z}_t = (\boldsymbol{d}_{1,t}, \ldots, \boldsymbol{d}_{N,t})'$. It follows for this reformulation that $\boldsymbol{\alpha}_t = \boldsymbol{\xi}_t$, $\boldsymbol{H} = \boldsymbol{\Phi}$, $\boldsymbol{\eta}_t = \boldsymbol{\zeta}_t$ and $\boldsymbol{\Sigma}_\zeta = \boldsymbol{\Sigma}_\eta$. In (1.8) we have merely adjusted the notation for expositional purposes under the temporary assumption that $\boldsymbol{\mu}_i = \mathbf{0}$.

For the linear Gaussian state space model (1.8) the likelihood can be evaluated by the prediction error decomposition that is provided by the Kalman filter; see Durbin & Koopman (2012, Chapter 7). In particular, the log likelihood is given by

$$\log(\boldsymbol{y}; \boldsymbol{\psi}) = \omega - \frac{1}{2}\sum_{t=1}^{T}\left(\log|\boldsymbol{F}_t| + \boldsymbol{v}_t'\boldsymbol{F}_t^{-1}\boldsymbol{v}_t\right), \tag{1.9}$$

where the prediction errors $\boldsymbol{v}_t$ and the prediction error variances $\boldsymbol{F}_t$ are computed by the Kalman filter that is given by

$$\begin{aligned} \boldsymbol{v}_t &= \boldsymbol{y}_t - \boldsymbol{X}_t\boldsymbol{\beta} - \boldsymbol{Z}_t\boldsymbol{a}_t, & \boldsymbol{F}_t &= \boldsymbol{Z}_t\boldsymbol{P}_t\boldsymbol{Z}_t' + \sigma^2\boldsymbol{I}_N, \\ \boldsymbol{a}_{t+1} &= \boldsymbol{H}\boldsymbol{a}_{t|t}, & \boldsymbol{P}_t &= \boldsymbol{H}\boldsymbol{P}_{t|t}\boldsymbol{H}' + \Sigma_\eta, \\ \boldsymbol{a}_{t|t} &= \boldsymbol{a}_t + \boldsymbol{P}_t\boldsymbol{Z}_t'\boldsymbol{F}_t^{-1}\boldsymbol{v}_t, & \boldsymbol{P}_{t|t} &= \boldsymbol{P}_t - \boldsymbol{P}_t\boldsymbol{Z}_t'\boldsymbol{F}_t^{-1}\boldsymbol{Z}_t\boldsymbol{P}_t, \end{aligned} \tag{1.10}$$

where $\boldsymbol{a}_t = \mathrm{E}(\boldsymbol{\alpha}_t|\boldsymbol{y}_1,\ldots,\boldsymbol{y}_{t-1};\boldsymbol{\psi})$ and $\boldsymbol{a}_{t|t} = \mathrm{E}(\boldsymbol{\alpha}_t|\boldsymbol{y}_1,\ldots,\boldsymbol{y}_t;\boldsymbol{\psi})$, are the predictive and filtered estimates for $\boldsymbol{\alpha}_t$. The corresponding variances matrices are given by $\boldsymbol{P}_t = \mathrm{Var}(\boldsymbol{\alpha}_t|\boldsymbol{y}_1,\ldots,\boldsymbol{y}_{t-1};\boldsymbol{\psi})$ and $\boldsymbol{P}_{t|t} = \mathrm{Var}(\boldsymbol{\alpha}_t|\boldsymbol{y}_1,\ldots,\boldsymbol{y}_t;\boldsymbol{\psi})$. The log likelihood (1.9) can be optimized with respect to the parameter vector $\boldsymbol{\psi}$ using the methods discussed in Durbin & Koopman (2012, Chapter 7).

When $\boldsymbol{\mu}_i$ is not equal to zero, the model (1.5) with $\boldsymbol{\xi}_t$ as in (1.7) can be referred to as a linear Gaussian state space model with random individual-specific effects. This model is not often used in panel data modeling nor state space analysis. Typically, state space approaches treat the individual-specific effects as fixed parameters, whereas random effects panel data approaches do not consider dynamic specifications for the time-varying effects $\boldsymbol{\xi}_t$. The advantages of having both random individual-specific effects and persistent random time-varying effects are efficiency gains and improved out-of-sample forecasting ability.

Model (1.5) with both random individual-specific and persistent time-varying random effects can still be cast in state space form. In particular, we may redefine the state space model in (1.8) where now $\boldsymbol{\alpha}_t = (\boldsymbol{\xi}_t', \boldsymbol{\mu}_1', \ldots, \boldsymbol{\mu}_N')'$ and the system matrices are given by

$$\boldsymbol{Z}_t = \begin{bmatrix} \boldsymbol{d}_{1,t}' & \boldsymbol{c}_{1,t}' & \boldsymbol{0} & \ldots & \boldsymbol{0} \\ \boldsymbol{d}_{2,t}' & \boldsymbol{0} & \boldsymbol{c}_{2,t}' & \ldots & \vdots \\ \vdots & & \ddots & & \boldsymbol{0} \\ \boldsymbol{d}_{N,t}' & \boldsymbol{0} & \ldots & \boldsymbol{0} & \boldsymbol{c}_{N,t}' \end{bmatrix}, \qquad \boldsymbol{H} = \begin{bmatrix} \boldsymbol{\Phi} & \boldsymbol{0} & \ldots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_q & & \vdots \\ \vdots & \vdots & \ddots & \boldsymbol{0} \\ \boldsymbol{0} & \ldots & \boldsymbol{0} & \boldsymbol{I}_q \end{bmatrix}$$

and

$$\boldsymbol{\Sigma}_\eta = \begin{bmatrix} \boldsymbol{\Sigma}_\zeta & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & & \vdots \\ \vdots & & \ddots & \mathbf{0} \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \end{bmatrix}.$$

Given the assumptions for the individual-specific and time-varying effects the distribution of the first state $\boldsymbol{\alpha}_1$ is normal with mean $\boldsymbol{a}_1 = \mathbf{0}$ and variance matrix $\boldsymbol{P}_1$ given by

$$\boldsymbol{P}_1 = \begin{bmatrix} \boldsymbol{I}_r & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_\mu & & \vdots \\ \vdots & & \ddots & \mathbf{0} \\ \mathbf{0} & \dots & \mathbf{0} & \boldsymbol{\Sigma}_\mu \end{bmatrix}.$$

The model is again a linear Gaussian state space model, where now the state $\boldsymbol{\alpha}_t$ includes both the time-varying effects and the individual-specific effects. The state vector is of length $Nq + r$. We notice that the individual-specific effects are not subjected to innovations after the first time period.

In principal, the Kalman filter can be used to evaluate the log likelihood for model (1.5). However, when the cross-section dimension N becomes large the Kalman filter becomes computationally demanding since the dimension of the state equation is a function of N. A straightforward computation of the prediction error decomposition for obtaining the likelihood might not be the most computationally efficient way for evaluating the likelihood.

In this thesis we are interested in deviations from the linear model in (1.5) with time-varying effects specified as in (1.7). As a special case, we develop new methods for the linear model (1.5). In particular, while the methods of Chapters 2 and 3 are developed for more general models, they apply to model (1.5).

1.2 Nonlinear Panel Data Models

The linear panel data model in (1.5) is quite restrictive for many empirical applications. We discuss four commonly encountered deviations from the linear model that are considered in this thesis.

First, the observations may follow non-Gaussian distributions. For example, suppose that we only observe the values $y_{i,t} = 0$ and $y_{i,t} = 1$, where $y_{i,t} = 1$ indicates some form of success. Clearly, in these situations modeling the observations by a Binary distribution is more appropriate than considering a Gaussian distribution. Also, count data and heavy-tailed data are more accurately modeled by non-Gaussian distributions. In general we model deviations from normality by replacing the observation equation (1.5) by

$$y_{i,t} \sim p(y_{i,t}|z_{i,t};\boldsymbol{\psi}), \qquad z_{i,t} = \boldsymbol{x}'_{i,t}\boldsymbol{\beta} + \boldsymbol{c}'_{i,t}\boldsymbol{\mu}_i + \boldsymbol{d}'_{i,t}\boldsymbol{\xi}_t, \tag{1.11}$$

where $p(y_{i,t}|z_{i,t};\boldsymbol{\psi})$ denotes an arbitrary non-Gaussian density that is fully determined by the signal $z_{i,t}$ and the parameter vector $\boldsymbol{\psi}$. The signal $z_{i,t}$ depends on a linear combination of explanatory variables, individual-specific and time-varying effects. The linearity assumption for $z_{i,t}$ is retained throughout this thesis. The linear Gaussian observation model in (1.5) is obtained as a special case when we choose $p(y_{i,t}|z_{i,t};\boldsymbol{\psi}) \equiv NID(z_{i,t}, \sigma^2)$

Many different choices for $p(y_{i,t}|z_{i,t};\boldsymbol{\psi})$ can be considered. Apart from Binary distributions, Poisson distributions and Student's t distributions have found wide applicability in economics and social sciences. The generalized panel data model (1.11) is the topic of Chapter 2.

Second, we obtain a nonlinear panel data model when the latent stochastic variables $\boldsymbol{\mu}_i$ and $\boldsymbol{\xi}_t$ are not considered additive. In particular, consider the following generalization of model (1.5)

$$y_{i,t} = \boldsymbol{x}'_{i,t}\boldsymbol{\beta} + h(\boldsymbol{\mu}_i, \boldsymbol{\xi}_t) + \epsilon_{i,t}, \qquad \epsilon_{i,t} \sim N(0, \sigma^2), \tag{1.12}$$

where $h(\cdot,\cdot)$ is an arbitrary function that maps the individual-specific and time-varying effects to the observations. A commonly encountered specification that applies when $q = r$ is given by

$$h(\boldsymbol{\mu}_i, \boldsymbol{\xi}_t) = \boldsymbol{\mu}'_i\boldsymbol{\xi}_t. \tag{1.13}$$

The function in (1.13) with $\boldsymbol{\xi}_t$ as in (1.7) is commonly referred to as a dynamic factor structure. Motivations for this type of structure are widespread. For example, when the observations $y_{i,t}$ refer to people the interaction between the individual-specific effects and the time-varying effects allows the individual-specific effects to vary stochastically over time. This is appropriate when arguing that the effect of the individual characteristics depends on the time period.

Further, from a macroeconomic perspective the model in (1.12) with $h(\boldsymbol{\mu}_i, \boldsymbol{\xi}_t)$ as in (1.13) is commonly referred to as the dynamic factor model; see Stock & Watson (2011). Dynamic factor models are used in macroeconomics for forecasting and structural analysis. The model imposes a dimension reduction when $r < N$, which is attractive for decomposing the variance of the observations into a common part and an idiosyncratic part. In Chapter 3 we consider the nonlinear model (1.12) with $h(\boldsymbol{\mu}_i, \boldsymbol{\xi}_t)$ as in (1.13). We show that this model specification, with both $\boldsymbol{\mu}_i$ and $\boldsymbol{\xi}_t$ stochastic, leads to an attractive model for forecasting in situations where there are many possible predictors.

Third, the random effects $\boldsymbol{\mu}_i$ and $\boldsymbol{\xi}_t$ may not only appear in the mean of the equation (1.5) but also in the variance. For example, in financial applications where the observations pertain to market returns, it is well known that the second moment of the distribution of the observations varies over time; see Engle (1982) and Shephard (2005). We may extend the specification of the error term in model (1.5) to obtain

$$\epsilon_{i,t} \sim N(0, \sigma_{i,t}^2), \qquad \sigma_{i,t}^2 = \exp(\tilde{\boldsymbol{d}}_{i,t}' \tilde{\boldsymbol{\xi}}_t), \tag{1.14}$$

where $\tilde{\boldsymbol{\xi}}_t$ follows a similar vector autoregressive process as given in (1.7). The exponential transformation ensures that the variance remains positive and at the same time creates the nonlinearity in the model. We have explicitly not included the individual-specific effects in (1.14) since we do not considered this situation in this thesis. A similar model with time-varying variances (1.14) is considered from a Bayesian perspective in Chib et al. (2006).

In Chapter 3 we consider models with both non-Gaussian features and time-varying means and variances. More specifically, we model the observations by the Student's t density, where both the mean and the variance are modeled by latent time-varying stochastic variables.

Fourth and finally, we consider the case where the latent time-varying effects follow a long-memory process. The presence of long-memory in a time series becomes apparent when its autocovariance function decays slower than an exponential decay. The time series is then said to be subject to long-range dependence; see Robinson (1994). A popular model specification for long-range dependent time series is given by the autoregressive fractionally integrated moving average (ARFIMA) model. For this model the elements of $\boldsymbol{\xi}_t$ are given by

$$\phi_j(B)\,(1-B)^{d_j}\,\xi_{j,t} = \theta_i(B)\zeta_{j,t}, \qquad \zeta_{ji,t} \sim \mathrm{N}(0, \sigma^2_{j,\zeta}), \tag{1.15}$$

for $j = 1, \ldots, r$, where B is the backshift operator for time index t with $B^m \xi_{j,t} = \xi_{j,t-m}$ for any integer m, the autoregressive $\phi_j(B)$ and moving average $\theta_j(B)$ are finite backshift polynomial functions and d_i is the fractional integration coefficient.

While this is not an entirely nonlinear model specification for the time-varying effects, it does prevent us from writing the model (1.5) in the finite state space form. Therefore the evaluation of the likelihood in (1.9) is not possible via the Kalman filter. In Chapter 4 of this dissertation we allow for general non-Gaussian model densities for the observations where the moments of the non-Gaussian distributions depend on the long-memory model (1.15) for the time-varying effects.

1.3 Why new methods are necessary

When any of the deviations from the linear Gaussian panel data model that are discussed above occur we cannot express the likelihood in closed form. To illustrate this consider the non-Gaussian panel data model in (1.11) with $\boldsymbol{\xi}_t$ as in (1.7). First, suppose that $\boldsymbol{\mu}_i = \mathbf{0}$, for all $i = 1, \ldots, N$. The resulting model falls into the class of non-Gaussian state space models; see Durbin & Koopman (2012, Part 2). The joint density of the observations can be expressed as a high dimensional integral by writing

$$\begin{aligned} p(\boldsymbol{y}; \boldsymbol{\psi}) &= \int_{\boldsymbol{\xi}} p(\boldsymbol{y}, \boldsymbol{\xi}; \boldsymbol{\psi}) \;\mathrm{d}\boldsymbol{\xi} \\ &= \int_{\boldsymbol{\xi}} p(\boldsymbol{y}|\boldsymbol{\xi}; \boldsymbol{\psi}) p(\boldsymbol{\xi}; \boldsymbol{\psi}) \;\mathrm{d}\boldsymbol{\xi}, \end{aligned} \tag{1.16}$$

where $\boldsymbol{\xi} = (\boldsymbol{\xi}_1', \ldots, \boldsymbol{\xi}_T')'$ and the integral is defined over all the possible paths of $\boldsymbol{\xi}$. When $p(\boldsymbol{y}|\boldsymbol{\xi}; \boldsymbol{\psi})$ is linear and Gaussian as in (1.8), the likelihood can be evaluated by the Kalman filter. For most other choices for $p(\boldsymbol{y}|\boldsymbol{\xi}; \boldsymbol{\psi})$ no closed from solution exists.

A simple solution for the evaluation of the integral in (1.16) can be obtained by Monte Carlo simulation. The idea is to view the integral as an expectation and to compute a Monte Carlo estimate for this expectation. In particular, such an estimate is given by

$$\hat{p}(\boldsymbol{y};\boldsymbol{\psi}) = M^{-1}\sum_{i=1}^{M} p(\boldsymbol{y}|\boldsymbol{\xi}^{(i)};\boldsymbol{\psi}), \tag{1.17}$$

where the samples $\boldsymbol{\xi}^{(i)}$ are drawn from $p(\boldsymbol{\xi};\boldsymbol{\psi})$. From the law of large numbers it follows that $\hat{p}(\boldsymbol{y};\boldsymbol{\psi})$ converges to $p(\boldsymbol{y};\boldsymbol{\psi})$ as $M \to \infty$. However, convergence will be slow even for small panel dimensions as the density $p(\boldsymbol{\xi};\boldsymbol{\psi})$ does not account for the observations $\boldsymbol{y}$. For empirically relevant panel sizes this approach is infeasible.

To obtain faster convergence rates for approximating the high dimensional integral in (1.16) different, more advanced, simulation methods have been developed. Examples, include importance sampling, Markov Chain Monte Carlo and particle filtering methods. These methods have been thoroughly developed for models with low cross-section dimensions and are extensively discussed in the textbooks of Doucet, de Freitas & Gordon (2001), Cappé, Moulines & Rydén (2005) and Durbin & Koopman (2012, Part 2). Next, we briefly discuss the importance sampling approach in some detail since it is of central importance in this thesis.

The importance sampling method proposes to rewrite the integral in (1.16) as follows

$$p(\boldsymbol{y};\boldsymbol{\psi}) = \int_{\boldsymbol{\xi}} \frac{p(\boldsymbol{y}|\boldsymbol{\xi};\boldsymbol{\psi})p(\boldsymbol{\xi};\boldsymbol{\psi})}{g(\boldsymbol{\xi}|\boldsymbol{y})} g(\boldsymbol{\xi}|\boldsymbol{y})\ \mathrm{d}\boldsymbol{\xi}, \tag{1.18}$$

where $g(\boldsymbol{\xi}|\boldsymbol{y})$ is the importance density. Now, instead of sampling from $p(\boldsymbol{\xi};\boldsymbol{\psi})$ we sample from $g(\boldsymbol{\xi}|\boldsymbol{y})$ and compute the Monte Carlo average

$$\hat{p}(\boldsymbol{y};\boldsymbol{\psi}) = M^{-1}\sum_{i=1}^{M} w^{(i)}, \qquad w^{(i)} = \frac{p(\boldsymbol{y}|\boldsymbol{\xi}^{(i)};\boldsymbol{\psi})p(\boldsymbol{\xi}^{(i)};\boldsymbol{\psi})}{g(\boldsymbol{\xi}^{(i)}|\boldsymbol{y})} \tag{1.19}$$

where the samples $\boldsymbol{\xi}^{(i)}$ are drawn from $g(\boldsymbol{\xi}|\boldsymbol{y})$ and the variables $w^{(i)}$ are the importance sampling weights, for $i = 1,\ldots,M$. In general the importance density $g(\boldsymbol{\xi}|\boldsymbol{y})$ must be close to the joint density $p(\boldsymbol{y},\boldsymbol{\xi};\boldsymbol{\psi})$, easy to sample from and easy to compute. The requirement for proportionality is formalized in Geweke (1989), who shows that if the variance of the importance sampling weights $w^{(i)}$ is finite the Monte Carlo estimate

$\hat{p}(\boldsymbol{y};\boldsymbol{\psi})$ converges to $p(\boldsymbol{y};\boldsymbol{\psi})$ with a $\sqrt{M}$ convergence rate.

Different methods for constructing the importance density are available; see Shephard & Pitt (1997), Durbin & Koopman (1997), Richard & Zhang (2007) and Koopman, Lucas & Scharth (2011) for examples. These methods are applicable in situations where the conditional density for the observations depends on either time-varying or individual-specific random effects. Also, the time-varying effects must expressible in a finite dimensional state space form. All the deviations that are discussed in the previous section imply that the standard importance sampling approach cannot be implemented without further modifications.

For the non-Gaussian panel data model with individual-specific and time-varying random effects in (1.11) we need to integrate both the individual-specific and time-varying effects from the joint density of the observations and the random effects. In particular, when $\boldsymbol{\mu}_i \neq \mathbf{0}$ the likelihood is given by

$$\begin{aligned} p(\boldsymbol{y};\boldsymbol{\psi}) &= \int_{\boldsymbol{\xi}}\int_{\boldsymbol{\mu}} p(\boldsymbol{y},\boldsymbol{\mu},\boldsymbol{\xi};\boldsymbol{\psi})\ \mathrm{d}\boldsymbol{\mu}\ \mathrm{d}\boldsymbol{\xi} \\ &= \int_{\boldsymbol{\xi}}\int_{\boldsymbol{\mu}} p(\boldsymbol{y}|\boldsymbol{\mu},\boldsymbol{\xi};\boldsymbol{\psi})p(\boldsymbol{\mu},\boldsymbol{\xi};\boldsymbol{\psi})\ \mathrm{d}\boldsymbol{\mu}\ \mathrm{d}\boldsymbol{\xi}, \end{aligned} \tag{1.20}$$

which is now a double integral over both the individual-specific and the time-varying effects. A standard importance sampling approach would rewrite the integral to obtain

$$p(\boldsymbol{y};\boldsymbol{\psi}) = \int_{\boldsymbol{\xi}}\int_{\boldsymbol{\mu}} \frac{p(\boldsymbol{y}|\boldsymbol{\mu},\boldsymbol{\xi};\boldsymbol{\psi})p(\boldsymbol{\mu},\boldsymbol{\xi};\boldsymbol{\psi})}{g(\boldsymbol{\mu},\boldsymbol{\xi}|\boldsymbol{y})} g(\boldsymbol{\mu},\boldsymbol{\xi}|\boldsymbol{y})\ \mathrm{d}\boldsymbol{\mu}\ \mathrm{d}\boldsymbol{\xi}, \tag{1.21}$$

where $g(\boldsymbol{\mu},\boldsymbol{\xi}|\boldsymbol{y})$ is the importance density.

For any choice of the density $g(\boldsymbol{\mu},\boldsymbol{\xi}|\boldsymbol{y})$ which accounts for the complicated a posteriori covariance structure of $\boldsymbol{\mu}$ and $\boldsymbol{\xi}$, sampling from it is likely to be complicated as the covariance matrix of $\boldsymbol{y}$ inherits the complicated structure. More specific, as $\boldsymbol{y}$ depends on both $\boldsymbol{\mu}$ and $\boldsymbol{\xi}$, there exists correlation between all individuals (due to $\boldsymbol{\xi}$) and time periods (due to $\boldsymbol{\mu}$). This makes sampling from $g(\boldsymbol{\mu},\boldsymbol{\xi}|\boldsymbol{y})$ infeasible even for moderate panel sizes. In Chapter 2 we show how to construct feasible importance densities for the non-Gaussian panel data model with individual-specific and time-varying random effects.

When the individual-specific and time-varying effects interact as in (1.12) with

$h(\boldsymbol{\mu}, \boldsymbol{\xi})$ as in (1.13), the construction of an adequate importance density is even more complicated. This is because small changes for $\boldsymbol{\mu}_i$ ($\boldsymbol{\xi}_t$) in $\boldsymbol{\mu}_i'\boldsymbol{\xi}_t$ have large impacts on the conditional mean and variance of $\boldsymbol{\xi}_t$ ($\boldsymbol{\mu}_i$). However, when the disturbances $\epsilon_{i,t}$ follow Gaussian distributions different methods can be used. In particular, we notice that when $\boldsymbol{\mu}_i$ is fixed at some value in (1.13), for $i = 1, \ldots, N$, the model (1.12) reduces to a linear Gaussian state space model. Also, when $\boldsymbol{\xi}_t$ is fixed, for $t = 1, \ldots, T$, the model reduces to a multivariate linear regression model. In Chapter 3 we exploit these aspects to develop an expectation-maximization algorithm for the computation of the posterior modes of $\boldsymbol{\mu}$ and $\boldsymbol{\xi}$. Also, additional importance sampling methods are developed for the evaluation of other posterior statistics. The importance sampling methods take advantage of the fact that the model reduces to a linear Gaussian state space model when the individual-specific effects are given.

When both the mean and variance of the panel data model depend on latent stochastic processes the standard importance sampling approach also becomes infeasible. To see this consider the linear panel data model in (1.5) with $\boldsymbol{\mu}_i = \mathbf{0}$ and the error term specified as in (1.14). The standard importance sampling specification for the likelihood is given by

$$p(\boldsymbol{y}; \boldsymbol{\psi}) = \int_{\boldsymbol{\xi}} \int_{\tilde{\boldsymbol{\xi}}} \frac{p(\boldsymbol{y}|\tilde{\boldsymbol{\xi}}, \boldsymbol{\xi}; \boldsymbol{\psi}) p(\tilde{\boldsymbol{\xi}}, \boldsymbol{\xi}; \boldsymbol{\psi})}{g(\tilde{\boldsymbol{\xi}}, \boldsymbol{\xi}|\boldsymbol{y})} g(\tilde{\boldsymbol{\xi}}, \boldsymbol{\xi}|\boldsymbol{y}) \, \mathrm{d}\tilde{\boldsymbol{\xi}} \, \mathrm{d}\boldsymbol{\xi}, \tag{1.22}$$

where $\boldsymbol{\xi}$ are the time-varying effects for the mean and $\tilde{\boldsymbol{\xi}}$ are the time-varying effects for the variance. The main difficulty is that both random effects are persistent time-varying stochastic processes. This complicates the construction of an adequate importance density $g(\tilde{\boldsymbol{\xi}}, \boldsymbol{\xi}|\boldsymbol{y})$. When the disturbances follow a Gaussian distribution we can exploit the fact that given $\tilde{\boldsymbol{\xi}}$ the model reduces to a linear Gaussian state space model; see Koopman & Bos (2004). This implies that only the integral over $\tilde{\boldsymbol{\xi}}$ needs to be approximated by Monte Carlo methods and that the integral over $\boldsymbol{\xi}$ can be calculated analytically given $\tilde{\boldsymbol{\xi}}$. However, when the observations are modeled by heavy-tailed distributions this is no longer possible. In Chapter 4 we develop adequate importance densities for an extended model with non-Gaussian observations and both stochastically time-varying means and variances. From a Bayesian perspective similar methods have been developed in Chib et al. (2006).

Finally, when the time-varying effects are modeled by the ARFIMA model in (1.15) the likelihood of the linear Gaussian state space model (1.9) can no longer be evaluated using the Kalman filter. This follows as the ARFIMA model cannot be expressed in a finite state space form. When the observations also follow non-Gaussian distributions the model becomes even more complicated. In Chapter 5 we present an importance sampling approach were both the non-Gaussian density and the long-memory processes are simultaneously approximated by a finite dimensional linear Gaussian state space model. These modifications for the importance sampling approach allow us to study non-Gaussian models with both latent long-memory and short-memory dynamics.

1.4 Empirical relevance

Panel datasets are commonly encountered in many research areas. Moreover, continuing efforts are made to collect new panels and to extend the current panels. Eventually most of the methods for panel data models will have to deal with both large cross-section and time series dimensions. Many of the datasets encountered may be modeled by nonlinear panel data models. We discuss three broad research areas for which the deviations from the linear model that are discussed above provide improvements over the linear model specification.

In micro-econometrics interest is often in individual-level panel data, where the time series pertain to people. The variables indicate different outcomes for different individuals. Popular panels include the Panel Study of Income Dynamics (PSID) and the Research and Development panel that is studied in for example Hausman, Hall & Griliches (1984). In this thesis we consider panels for union membership and criminal convictions that were previously studied in Vella & Verbeek (1998) and Bijleveld & Wijkman (2009).

These panels have in common that many of the variables that are recorded are non-Gaussian in nature. Count variables and indicator variables are common. Despite the many nonlinearities, the objective of conducting valid inference on the model parameters remains the same. Also, at the same time selection effects, or unobserved variables are still likely to be present. This requires the separation of observed and unobserved effects. Therefore flexible panels that allow for both individual-specific and time-varying effects are of vital importance.

In macroeconomics large numbers of time series are available that describe various components of the economy. The contributions of Stock & Watson (2002*a*) and Stock & Watson (2002*b*) have started a comprehensive literature that aims to combine these large quantities of available time series in order to obtain reductions in the forecast errors. Additionally, controlling for unobserved heterogeneity in vector autoregressive models by appending factor structures has found widespread applicability in macroeconomics; see Stock & Watson (2005) and Bernanke, Boivin & Eliasz (2005).

Typically, the factor decomposition is based on fixed parameters. The individual-specific loadings which relate the common time-varying effects to the observations are treated as fixed parameters. In this thesis we show that by treating the individual-specific loadings and the common time-varying factors as latent stochastic processes the forecast errors can be greatly reduced. Additionally, we show that the Student's t distribution allows for more accuracy when extracting the individual-specific and time-varying effects from macroeconomic time series. In the sense that the confidence bounds are smaller. This is relevant for panels of economic growth, such as the Penn World Tables, where many outliers are present which can have large effects on the parameter estimates; see De Long & Summers (1991).

In finance many different panels are collected. In this thesis we consider panels from stock market returns and panels of interest rates. The characteristics of the financial time series are more complicated due to the high sampling frequency. This causes correlation in the second moments of the observed variables. We contribute to the literature by providing classical methods for the simultaneous analysis of multiple time series with both stochastically time-varying means and variances, non-Gaussian distributions for the observations, and possible long-memory features.

1.5 Thesis overview

The core of this thesis contains five self-contained chapters. For expositional purposes the notation is adjusted in each Chapter to satisfy the objectives of the chapter.

In the second chapter we develop an exact maximum likelihood method for the estimation of parameters in a nonlinear non-Gaussian dynamic panel data model with unobserved random individual-specific and time-varying effects. We propose an estimation procedure based on the importance sampling technique. In particular, a sequence of conditional importance densities is derived which integrates out all random effects from the joint distribution of endogenous variables. We disentangle the integration over both the cross-section and the time series dimensions. The estimation method facilitates the modeling of large panels in both dimensions. We evaluate the method in an extended Monte Carlo study for dynamic panel data models with observations from different non-Gaussian distributions. We finally present three empirical illustrations for (i) union choice of young males using a Binary panel, (ii) crime rates of families using a Binomial panel and (iii) economic growth modeling using a Student's t panel.

In the third chapter we consider the dynamic factor model for the case where the rows of the loading matrix, the dynamic factors and the disturbances are treated as latent stochastic processes. We present parametric empirical Bayes methods that enable the efficient estimation of the loadings and the factors. First, we provide an iterative filtering algorithm to evaluate the joint posterior mode of the loadings and the factors. Second, we show that the vector of deterministic model parameters can be estimated using a two-step likelihood-based approach. Third, we show that other posterior statistics for the loadings and the factors, such as the posterior mean and variance, can be calculated using basic simulation methods. By treating the model components stochastic and estimating the parameters using likelihood-based methods we obtain efficient shrinkage-type estimates for both the loadings and the factors. We show that the joint posterior mode estimates for both the loadings and the factors have lower quadratic loss compared to the standard maximum likelihood estimates. We investigate the methods in a Monte Carlo study where we document the finite sample properties. Further, the methods are illustrated for the forecasting of macroeconomic and financial time series as well as decomposing economic growth rates.

In the fourth chapter, we develop methodology for the joint modeling of the yield curve and non-standard monetary policy during crisis times. The yield curve is modeled by the three factor dynamic Nelson-Siegel model and the monetary policy measurements are modeled by appropriate non-Gaussian densities that are defined conditional on a set of latent dynamic factors. The term structure and monetary policy factors are jointly updated by a latent vector autoregressive process. Further, to capture outliers and changing levels of volatility for the yields, we extend the specification of the yield curve model to allow for heavy tailed errors and common factor stochastic volatility. The complete model is non-Gaussian and has both time-varying means and time-varying variances. We develop an estimation procedure that is based on a novel implementation of the importance sampling technique to estimate the model parameters. We disentangle the integration over both the time-varying means and the time-varying variances. To illustrate the model we consider the yields from four European countries and the monetary policy of the European Central Bank. We model the European short rate by the log-normal distribution to account for the zero-lower bound and we model the bond market purchases that were conducted by the European Central Bank under the Securities Markets Programme by the Poisson distribution. We document the interaction between the yield curves and the monetary policy using parameter estimates and impulse response functions.

In the fifth chapter an exact maximum likelihood method is developed for the estimation of parameters in a non-Gaussian nonlinear density function that depends on a latent Gaussian dynamic process with long-memory properties. Our method relies on the method of importance sampling and on a linear Gaussian approximating model from which the latent process can be simulated. Given the presence of a latent long-memory process, we require a modification of the importance sampling technique. In particular, the long-memory process needs to be approximated by a finite dynamic linear process. Two possible approximations are discussed and are compared with each other. We show that an autoregression obtained from minimizing mean squared prediction errors leads to an effective and feasible method. In our empirical study we analyze ten log-return series from the S&P 500 stock index by univariate and multivariate long-memory stochastic volatility models.

In the sixth and final chapter we consider the one-step ahead forecasting of the outcome of the annual boat race between Cambridge and Oxford. Initially, we compare the relative performance of different dynamic models for forty years of forecasting. Each model is defined by a binary density conditional on a latent signal, which is modeled by a variety of dynamic stochastic components and fixed predictors. The out-of-sample predictive ability of the models is compared between each other by using a variety of loss functions and predictive ability tests. We find that the models where the latent signal is modeled by an AR(1) or stochastic cycle process cannot be outperformed by other models and are able to correctly predict 30 out of 40 races. Further, we investigate the sensitivity of our results to the chosen sample split point of forty years. We find that the forecasting performance and the predictive ability tests are highly sensitive to the chosen sample split point. Sample split robust predictive ability tests indicate that the models where the signal is modeled by an AR(1) or ARFI(0,d) process are most unlikely to be outperformed.