



TI 2007-027/4

Tinbergen Institute Discussion Paper

Estimating Systematic Continuous-time Trends in Recidivism using a Non-Gaussian Panel Data Model

Siem Jan Koopman¹

André Lucas¹

Marius Ooms¹

Kees van Montfort¹

Victor van der Geest²

¹ *Vrije Universiteit Amsterdam,*

² *Netherlands Institute for the Study of Crime and Law Enforcement (NCSR), Leiden.*

Tinbergen Institute

The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam, and Vrije Universiteit Amsterdam.

Tinbergen Institute Amsterdam

Roetersstraat 31

1018 WB Amsterdam

The Netherlands

Tel.: +31(0)20 551 3500

Fax: +31(0)20 551 3555

Tinbergen Institute Rotterdam

Burg. Oudlaan 50

3062 PA Rotterdam

The Netherlands

Tel.: +31(0)10 408 8900

Fax: +31(0)10 408 9031

Most TI discussion papers can be downloaded at
<http://www.tinbergen.nl>.

Estimating systematic continuous-time trends in recidivism using a non-Gaussian panel data model

Siem Jan Koopman^{*}, André Lucas^{**}, Marius Ooms^{*},
Kees van Montfort^{**}, Victor van der Geest^{***}

^{*} *Department of Econometrics, Vrije Universiteit*

^{**} *Department of Finance, Vrije Universiteit*

De Boelelaan 1105, NL-1081 HV Amsterdam

^{***} *Netherlands Institute for the Study of Crime and Law Enforcement (NCSR)*
P.O. Box 792, NL-2300 AT Leiden

March 7, 2007

Abstract

We model panel data of crime careers of juveniles from a Dutch Judicial Juvenile Institution. The data are decomposed into a systematic and an individual-specific component, of which the systematic component reflects the general time-varying conditions including the criminological climate. Within a model-based analysis, we treat (1) shared effects of each group with the same systematic conditions, (2) strongly non-Gaussian features of the individual time series, (3) unobserved common systematic conditions, (4) changing recidivism probabilities in continuous time, (5) missing observations. We adopt a non-Gaussian multivariate state space model that deals with all of these issues simultaneously. The parameters of the model are estimated by Monte Carlo maximum likelihood methods. This paper illustrates the methods empirically. We compare continuous-time trends and standard discrete-time stochastic trend specifications. We find interesting common time-variation in the recidivism behavior of the juveniles during a period of 13 years, while taking account of significant heterogeneity determined by personality characteristics and initial crime records.

Keywords: non-Gaussian state space modeling, nonlinear panel data model, binomial time series, recidivism behavior, continuous time modelling.

*Correspondence to: kvmontfort@feweb.vu.nl

1 Introduction

Becker (1968) is often used as a starting point to model recidivism behavior (see Corman and Mocan, 2000). According to this standard theoretical framework, the optimization of individuals engaged in criminal activities, depends upon the expected payoffs of the criminal activities. Instances are the return to legal labor-market activities and the costs of criminal activities, such as conviction and punishment. Some studies provided evidence that increases in criminal-justice sanctions reduce criminal activities (see Levitt, 1997), while others found either a weak relationship, or none at all between the two (see Cornwell and Trumbull, 1994). The analysis of official criminal trajectories of individuals is an important topic in the research field of recidivism behavior. Effort has been expended on examining the changing frequency of criminal activities over time, and distinguishing different criminal trajectories based on these frequencies. Among others, Nagin and Land (1993), Nagin et al. (1995), D’Unger et al. (1998), Roeder et al. (1999) and White et al. (2001) use semi-parametric Poisson models to handle these trajectories. Other research has also dealt with criminal trajectories of specific types of behavior, such as violent crime or serious offenders (see, among others, Brame et al., 2001, and Piquero et al., 2001). The aim of this research was to assess trajectory groups based on frequency, under the assumption that an offender’s membership stays fixed throughout the life course. Nevertheless, this assumption is rather restrictive and not realistic.

There are several articles that deal with the identification of types of criminal behavior and the interaction of these types (see Francis et al., 2004). Among others, Wiggins (1973), Stander et al. (1989), Bijleveld and Mooijaart (2003), and Bartolucci et al. (2007) use Markov Models. They assume that the offending pattern of a subject within a certain age period depends only on a discrete latent variable representing the tendency to commit crimes, which follows a first-order non homogeneous Markov Process. The model allows for the presence of discrete covariates affecting the initial and the transition probabilities of the latent processes.

In this article we adopt a dynamic statistical model for the intensity of recidivism as measured in criminal records of the juveniles. We model the probability of committing a crime in a certain calendar month using a non-Gaussian panel data model with multiple deterministic factors and dynamic random effects. In order to do this, we explore the persistence and existence of common factors in calendar time in the crime careers of juveniles from the Dutch correctional facility of Harreveld, a Judicial Juvenile Institution. Although not all juveniles had a criminal record before entering the facility, we use the common term recidivism. The common factors are modelled in continuous time and may partly be attributed to macro measures that influence the crime trickling down into this group. Since changes in these measures may not directly affect crime, we specify the factors as smooth continuous time processes. In our approach we further aim to distinguish between effects on the level of recidivism and on the persistence of recidivism.

The data allow us to distinguish age effects, cohort effects, personality characteristics and social factors, variables which all have been measured during the stay at the juvenile justice institution. Given the restricted number of individuals we limit ourselves here to one personality characteristic, one general individual risk factor and a common time-varying risk factor. Interesting personal characteristics are consciousness and intelligence. Generally important predictive factors are the individual criminal records before entering the juvenile justice institution and the criminal record of relatives. An economic factor is family unemployment. All these factors were found to be significant in a descriptive cross-section analysis of risk determining the average level of recidivism (i.e. number of crimes per month), the probability of excessive recidivism (extreme number of crimes per month) and the speed of recidivism (i.e. duration until next crime) in an earlier analysis of the data by Van Der Geest, Bijleveld and Wijkman (2005).

In the approach of this paper, the observed time series are decomposed into a systematic and an individual-specific component, of which the systematic component reflects the general community conditions and the criminological climate. We set ourselves to treat (1) shared effects of each group to the same systematic conditions, (2) strongly non-Gaussian features of the individual time series, (3) possible continuous time dynamics of an unobserved common systematic condition, (4) changing probabilities of recidivism over time, (5) missing observations. This is a challenging task and innovative within a criminological context. We adopt the non-Gaussian panel data model with latent dynamic factors of Koopman and Lucas (2007). They have used the model in a credit risk analysis to identify common and idiosyncratic risk factors for defaults of firms. Their non-Gaussian multivariate state space model can treat all issues that we require for the analysis of recidivism simultaneously. However, we extend their methodology by adopting continuous time processes for the unobserved common factors. Econometric continuous-time models were first developed by Bergstrom. Harvey (1989, Chapter 9) extended the continuous-time methodology to structural unobserved component models for trend-cycle decompositions of economic data. Singer (1998) developed continuous-time for time-varying parameter panel data models and Oud and Jansen (2000) discuss continuous-time state space modelling of panel data in a dynamic structural equations framework. As our panel consists of many points in time, we can allow for different time series specifications of the trend. We compare the continuous time model with the discrete time model of Koopman and Lucas (2007).

The parameters of the models are estimated using computationally efficient state space methods together with importance sampling techniques, see Durbin and Koopman (2001). Methodological and computational developments allow us to analyse such complex and high-dimensional data-sets in a fast and satisfactory way on a standard desk-top computer.

We show that the analysis of the Harreveld data-set using our modelling framework leads to a better understanding of the recidivism behavior of juveniles. For example, we have found

a significant downward common trend in the intensity of serious crime in our sample.

The outline of this paper is as follows. We start with a description of the model. Next we will discuss the Monte Carlo estimation methodology used in order to estimate parameters of the model. After that, we deal with the data and we present empirical results illustrating the relevance of our method for the data at hand. Finally, some conclusions are presented.

2 The model for crime intensities

The dynamic model describes the criminal behaviour of a specific group of juveniles in calendar time $t = 1, \dots, T$. The crime dynamics of these young adolescents are only observed after they leave a program organised by the justice department as an alternative to a jail sentence. The number of individuals grows in the beginning of the sample since the program started at the beginning of the sample. Nearly all individuals are observed until the end of our sample.

The variable of interest is the intensity of juvenile crime as measured by the proportion of juveniles committing offenses in a certain month of the year. This enables us to look at common effects in calendar time, as is usual in the economic time series analysis of crime, see e.g. Corman and Mocan (2000). In such analyses, trends, seasonals and structural changes in crime levels are the key properties to be modeled and explained by common demographic, economic and deterrence factors. Our approach examines whether a macro effect is important in explaining offense levels after taking individual characteristics into account. The macro effect is modeled as a continuous time unobserved component process and is common to all individuals in the observed group. In this way we can assess whether general macro effects are due to variations in juvenile crime or whether specific factors are more important. Policies for reducing juvenile crime can then be targeted more effectively.

We distinguish different classes of offenses, indexed by $s = 1, \dots, S$. We condition on a categorical background variable j for a personal characteristic and on an initial categorical criminal record variable a with $j = 1, \dots, J$ and $a = 1, \dots, A$. Our model uses the non-Gaussian panel time series specification developed by Koopman and Lucas (2007). The basic structure of this model is derived from a common framework in credit risk modelling. It decomposes the probability of a default of a firm into systematic and firm-specific risk components, see the influential framework of Gupton, Finger, and Bhatia (1997). In this paper we employ this framework to model crime data, reversing the suggestion of Schmidt and Witte (1989) to use ideas from statistical crime analysis to empirical bankruptcy modeling.

2.1 A dynamic limited dependent variable panel data model

The model specification builds on the well known limited dependent variable interpretation of discrete choice models. We consider a latent index variable I_{it} for each juvenile i at time t for $i =$

$1, \dots, N$ and $t = 1, \dots, T$. An offense is committed if the index variable falls below a threshold c_{it} . The threshold is individual-specific due to differences in the individual characteristics and personal situations determining crime. Although characteristics are not typically time-varying, personal situations can vary over time. To allow for dependence among the different individuals, we assume that the index variable $I_i(\tau)$ is driven by a continuous systematic risk factor $f(\tau)$ and an individual specific risk factor $x_i(\tau)$,

$$I_i(\tau) = \rho_i f(\tau) + \sqrt{1 - \rho_i^2} x_i(\tau), \quad i = 1, \dots, N, \quad 0 \leq \tau \leq n, \quad (1)$$

compare Gupton et al. (1997). In order to identify the model we assume that the individual-specific factors $x_i(\tau)$ are serially and mutually independent, conditional on the systematic factor $f(\tau)$. The parameter ρ_i in (1) determines the relative weights of the systematic and individual-specific risk component in the crime intensity. Its value is strictly between 0 and 1. For $\rho_i = 0$, crime is entirely triggered by individual-specific factors $x_i(\tau)$. By contrast, for $\rho_i = 1$, the variation in offenses is completely driven by the common factor $f(\tau)$. The weighting scheme in (1) ensures identification of the final model below. We should note that identification is an issue because $I_i(\tau)$ is not observed directly. The variances of $f(\tau)$ and $x_i(\tau)$ need to be restricted. In case the variances of $f(\tau)$ and $x_i(\tau)$ are restricted to be unity, the correlation between $I_i(\tau)$ and $I_{i'}(\tau)$ for $i \neq i'$ equals $\rho_i \rho_{i'}$, or ρ_i^2 for homogeneous entities $\rho_i = \rho_{i'}$.

Instead of focusing on individual crime intensities over longer periods of time, we concentrate on the aggregate offense level within subgroups of juveniles for relatively short periods of time. Let k_{sjat} denote the number of juveniles in offense class s , psychological category j , and initial criminal record class a , observed at period τ_t , and let y_{sjat} denote the number of offending individuals in the category s, j, a and for period τ_t for $t = 1, \dots, T$. Then y_{sjat} can be modelled as a realization of a binomial distribution. We do not observe exogenous time-varying covariates, but there are many reasons to believe that the offense intensity is time-varying for the juveniles under scrutiny. The most obvious reason is the age distribution of the juveniles which changes as time progresses in our sample. Also, economic conditions and government policies towards juvenile crime can change over time. We treat the time-variation as an unobserved component or latent factor $f(\tau)$. If the changing age distribution is the only time-effect causing the time-variation in the offense intensity, we could interpret $\hat{f}(\tau)$ as a flexible estimator of the functional form of the time effect of the changing age distribution.

The binomial variable y_{sjat} is observed at time point τ_t and modelled conditional on the systematic risk factor $f_t \equiv f(\tau_t)$, i.e.,

$$y_{sjat} | f_t \sim \text{Binomial}(k_{sjat}, \pi_{sjat}), \quad \text{with } f_t = f(\tau_t), \quad (2)$$

for $s = 1, \dots, S$, $j = 1, \dots, J$, $a = 1, \dots, A$, $t = 1, \dots, T$, where π_{sjat} is the conditional probability of committing a crime in category s, j, a and time τ_t .

Figure 1 shows plots of the proportions data for crimes of category $s = 1$, y_{2jat}/k_{2jat} , against time τ_t for $t = 1, \dots, T$. The discreteness of the data is clearly visible. A Gaussian modelling approach for this type of data should be rejected.

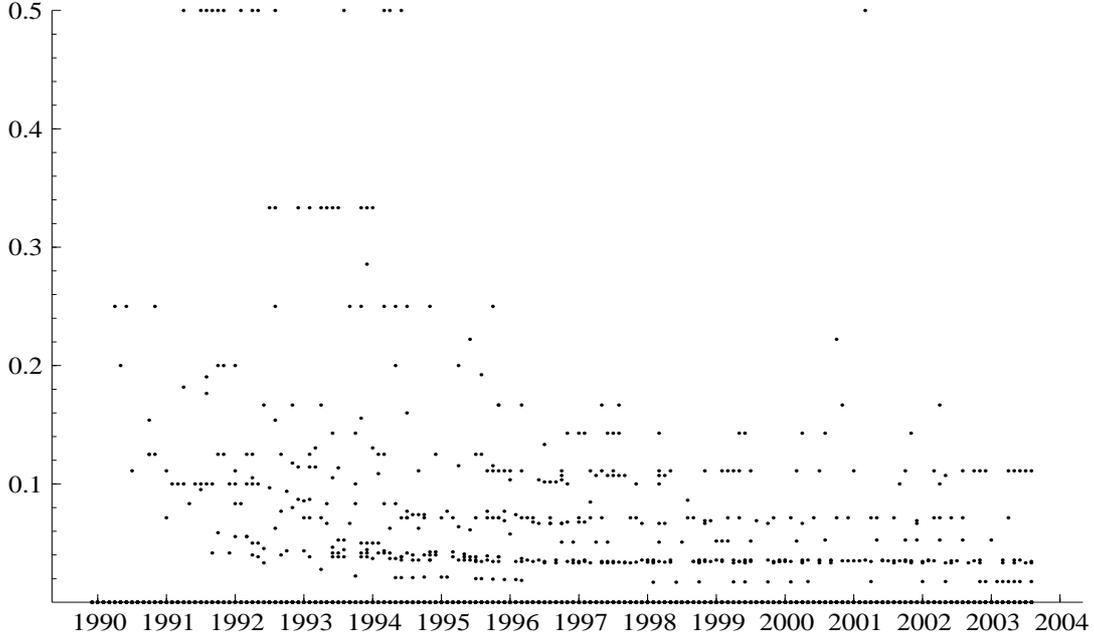


Figure 1: Category 2 crime intensity data y_{2jat}/k_{2jat} , $j = 1, 2, 3$, $a = 1, \dots, 4$, against time τ_t for $t = 1, \dots, 165$.

To reduce the number of parameters, we impose the following restrictions. First, we restrict the thresholds c_i to depend only on s , j , and a , i.e., $c_i = \bar{c}_{sija_i}$. In particular, we impose the additive structure

$$\bar{c}_{sja} = \bar{c}_0 + \bar{c}_{1,s} + \bar{c}_{2,j} + \bar{c}_{3,a}.$$

Dependence of different measures of crime intensity on personal characteristics, indexed by j , and the initial criminal record, indexed by a , was found to be significant in earlier studies of this data, see Van der Geest and Bijleveld (2006). In principle, we allow the correlation parameter ρ_i^2 to depend on crime category, on personal characteristics or initial crime record, i.e., $\rho_i = \bar{\rho}_{sija_i}$.

From (1) and the restrictions imposed, it follows that

$$\pi_{sjat} = \Pr \left(x_i(\tau) \leq \frac{\bar{c}_{sja} - \bar{\rho}_{sja} f t}{\sqrt{1 - \bar{\rho}_{sja}^2}} \right) = F \left(\frac{\bar{c}_{sja} - \bar{\rho}_{sja} f t}{\sqrt{1 - \bar{\rho}_{sja}^2}} \right) \quad (3)$$

where $F(\cdot)$ is a cumulative distribution function of the firm-specific risk factor. The probability of offense can now vary between offense categories s , personal characteristics a , initial crime records j and the time-varying risk factor $f(\tau)$.

The unobserved risk process $f(\tau)$ is evolving in continuous time. We adopt a continuous time trend model specification for the risk $f(\tau)$ with slope $g(\tau)$. A suitable model for the slope is the continuous process $dg(\tau) = \sigma_g dw(\tau)$ where $w(\tau)$ is standard Brownian motion with $\sigma_g > 0$. A smooth trend model in continuous time is then obtained by

$$f(\tau) = f(0) + \int_0^\tau g(r)dr = f(0) + g(0)\tau + \sigma_g \int_0^\tau w(r)dr. \quad (4)$$

The unknown coefficients for this model are the initial trend coefficient $f(0)$ and $g(0)$ together with the standard deviation σ_g . We impose a unit conditional variance restriction, $\sigma_g = 1$ on $f(\tau)$, because the standard deviation σ_g and the correlation $\bar{\rho}_{sja}^2$ cannot be identified simultaneously. Theoretical extensions to more general processes for $f(\tau)$ are straightforward in a continuous time state space framework, see Harvey (1989, Chapter 9). Also generalizations to models with multiple factors can be considered, see, e.g., Harvey (1989, Chapter 8).

Given that we have aggregated monthly data for approximately 13 years, we can also consider a discrete-time specification, where f_t follows a simple random walk where t is measured in months. This is one of the models considered by Koopman and Lucas (2007):

$$f_t = f_{t-1} + \eta_t, \quad t = 1, \dots, T, \quad (5)$$

where η_t is standard normally distributed and serially uncorrelated. In the empirical application, we estimate several extended models.

Explanatory variables can be added to the model such as macroeconomic variables. Also individual-specific covariates can be incorporated and more category types can be introduced in the model. However, given the limited amount of information in the data, the current three-way model with a limited amount of categories for the heterogeneity in individual characteristics seems appropriate. Adding more categories and having more covariates may makes it impossible to empirically identify all parameters. A sufficient amount of pooling is needed to obtain reasonable estimates.

2.2 A model reformulation for the statistical analysis

The unknown parameters of our crime model are \bar{c}_{sja} , and $\bar{\rho}_{sja}$. To estimate these parameters we develop a statistical model that allows feasible maximization likelihood estimation of the parameters. The statistical model also provides the means for estimating the latent process $f(\tau)$ (signal extraction). It is clear that standard linear estimation methods cannot be applied since the model is non-Gaussian and depends on a continuous time latent process $f(\tau)$. Monte Carlo simulation techniques are carefully designed and employed to carry out the task of estimation in a computationally efficient way. This is a challenging task since the dimensions of the integrals involved in dynamic limited dependent variables models are high. In the empirical section we consider a set of models introduced in Section 2.1 with $S = 1$, $J = 3$, $A = 4$ and $T = 165$

resulting in 165 observations of a 12×1 vector. We also allow for up to 4 different risk factors. All estimations proved computationally very feasible on ordinary desktop PCs.

We start with (2) and consider the conditional Binomial log-density function as given by

$$\log p(y_{sjat}|\pi_{sjat}) = y_{sjat}\{\log \pi_{sjat} - \log(1 - \pi_{sjat})\} + k_{sjat} \log(1 - \pi_{sjat}) + \log \binom{k_{sjat}}{y_{sjat}}, \quad (6)$$

with $y_{sjat} = 0, \dots, k_{sjat}$. As the probability π_{sjat} is bounded between zero and one, we introduce the logit transformation

$$\theta_{sjat} = \log\{\pi_{sjat} / (1 - \pi_{sjat})\} \quad (7)$$

such that $F(\cdot)$ in (3) is effectively the logistic distribution. This leads to the following convenient expression for the log-density of y_{sjat} ,

$$\log p(y_{sjat}|\theta_{sjat}) = y_{sjat}\theta_{sjat} - k_{sjat} \log(1 + \exp(\theta_{sjat})) + \log \binom{k_{sjat}}{y_{sjat}}, \quad (8)$$

from which it follows immediately that the Binomial density belongs to the exponential family. The model for the signal θ_{sjat} is specified as

$$\theta_{sjat} = \lambda_0 + \lambda_{1,s} + \lambda_{2,j} + \lambda_{3,a} - (\beta_0 + \beta_{1,j} + \beta_{2,a})f_t, \quad (9)$$

where $f_t = f(\tau_t)$ is the systematic risk factor of the index variable I_{it} in (1) at time τ_t . One of the $\lambda_{1,s}$ (and of the $\lambda_{2,j}$, $\lambda_{3,a}$, $\beta_{1,s}$, and $\beta_{2,j}$) is restricted to zero to avoid multicollinearity in equation (9). This statistical framework can be related to the model parameters in Section 2.1. For example, if $\bar{\rho}_{saj} = \bar{\rho}_{aj}$ one uses the relations

$$\begin{aligned} \bar{\rho}_{aj}^2 &= (\beta_0 + \beta_{1,a} + \beta_{2,j})^2 / (1 + (\beta_0 + \beta_{1,a} + \beta_{2,j})^2), \\ \bar{c}_{sja} &= (\lambda_0 + \lambda_{1,s} + \lambda_{2,j} + \lambda_{3,a}) / \sqrt{1 + (\beta_0 + \beta_{1,a} + \beta_{2,j})^2}. \end{aligned} \quad (10)$$

The latent factor $f_t \equiv f(\tau_t)$ in (9) is obtained from the unobserved process $f(\tau)$ that is modeled by the continuous time process (4). Other continuous time processes for $f(\tau)$ can be considered in this context as well.

Since the observations are recorded at discrete time intervals, we wish to formulate the continuous time process (4) by the discrete state space model

$$\begin{pmatrix} f_{t+1} \\ g_{t+1} \end{pmatrix} = \begin{bmatrix} 1 & \delta_t \\ 0 & 1 \end{bmatrix} \begin{pmatrix} f_t \\ g_t \end{pmatrix} + \begin{pmatrix} \eta_t \\ \zeta_t \end{pmatrix}, \quad t = 1, \dots, T, \quad (11)$$

where $f_t = f(\tau_t)$, $g_t = g(\tau_t)$ and $\delta_t = \tau_{t+1} - \tau_t$. Furthermore, we have

$$\eta_t = \sigma_g \int_0^{\tau_{t+1}} [w(r) - w(\tau_t)] dr, \quad \zeta_t = \sigma_g [w(\tau_{t+1}) - w(\tau_t)],$$

which follows from (4). It can be shown that these disturbances are normally and independently distributed (NID) with zero means while their variances and covariance are given by

$$\text{Var} \begin{pmatrix} \eta_t \\ \zeta_t \end{pmatrix} = \sigma_g^2 \delta_t \begin{bmatrix} \frac{1}{3} \delta_t^2 & \frac{1}{2} \delta_t \\ \frac{1}{2} \delta_t & 1 \end{bmatrix}, \quad (12)$$

see Durbin and Koopman (2001, §3.10.2). In our example δ_t could vary according to the length of the month, measured in days. Because we don't use the exact dates in calendar time, we simply take $\delta_t = 1$. We stress this is not a restriction imposed by the methodology. However, this restriction allows us to compare our results with the simple discrete time specification used by Koopman and Lucas (2007). The comparison of the two trend models illustrates the flexibility of non-Gaussian state space modelling with respect to dynamic specification.

The crime intensity model can be represented as a non-Gaussian state space panel data model with observation equation

$$y_{sjat} \sim p(y_{sjat} | \theta_{sjat}) = \text{Binomial}(k_{sjat}, \pi_{sjat}), \quad (13)$$

where $\pi_{sjat} = (1 + \exp(-\theta_{sjat}))^{-1}$ and θ_{sjat} is given by (9). The dynamic latent factor processes $f(\tau)$ and $g(\tau)$ together with the fixed effects λ_0 , $\lambda_{1,s}$, $\lambda_{2,j}$ and $\lambda_{3,a}$ can be estimated simultaneously in this methodology. We represent the latent factors in discrete time and place them, together with the fixed effects, in a so-called state vector α_t that is modeled by the linear time series process

$$\alpha_{t+1} = T_t \alpha_t + R_t \xi_t, \quad \xi_t \sim \text{NID}(0, Q_t), \quad (14)$$

for $t = 1, \dots, T$ with $\alpha_1 \sim N(\bar{\alpha}_0, P)$. The vector $\bar{\alpha}_0$ and the matrices T_t , R_t , Q_t and P are fixed and known matrices for $t = 1, \dots, T$. The signal is then given by

$$\theta_{sjat} = Z_{sjat} \alpha_t, \quad (15)$$

where Z_{sjat} is a deterministic selection vector for $s = 1, \dots, S$, $j = 1, \dots, J$, $a = 1, \dots, A$, and $t = 1, \dots, T$. Some elements of the system matrices T_t and Q_t and system vector Z_{sjat} can depend on unknown coefficients. These coefficients are collected in a parameter vector and their estimation is discussed in the next section.

To illustrate the state space representation of our model, we consider an example with $S = A = 1$ and $J = 3$ types of personal characteristics. The state vector becomes $\alpha_t = [\lambda_0, \lambda_{2,1}, \lambda_{2,2}, f_t, g_t]'$ where f_t and g_t are modelled as in (11). Further, we restrict $\lambda_{1,1} = \lambda_{2,3} = \lambda_{3,1} = 0$ for identification. In this example, the system matrices for the linear time series process are given by

$$T_t = \begin{bmatrix} I_3 & 0_{3 \times 1} & 0_{3 \times 1} \\ 0_{1 \times 3} & 1 & \delta_t \\ 0_{1 \times 3} & 0 & 1 \end{bmatrix}, \quad R_t = \begin{bmatrix} 0_{3 \times 2} \\ I_2 \end{bmatrix},$$

and Q_t is given by (12). The state space vectors for the signal are given by

$$(Z'_{1,1,1,t}, Z'_{1,2,1,t}, Z'_{1,3,1,t})' = (I_3, -(\beta_0 + \beta_1), 0_{3 \times 1}),$$

where $\beta'_1 = (\beta_{2,1}, \beta_{2,2}, \beta_{2,3})$ and $\beta_{2,3} = 0$ for identification.

The joint model (13) and (14) is a special case of a nonlinear and non-Gaussian state space model as discussed in part II of Durbin and Koopman (2001). In the next section we focus on the estimation of the parameters λ and β by Monte Carlo maximum likelihood. We should note that fixed effects, λ , are incorporated in the state vector α_t and therefore estimated jointly with latent factors f_t and g_t . The state vector is estimated as part of the Monte Carlo methods used for the evaluation of the likelihood function.

3 Monte Carlo estimation

The statistical modelling framework presented in the previous section is non-Gaussian and therefore standard time series methods such as the Kalman filter for the evaluation of the likelihood function can not be used. The likelihood function is no longer of a closed form for such general models and therefore can not be computed by analytical expressions. Simulation techniques have been developed for classical estimation but also for Bayesian estimation approaches. In this paper we restrict ourselves to classical estimation that is based on maximising the so-called Monte Carlo likelihood function.

3.1 Importance sampling

The model consists of a Binomial measurement equation (13) for $y = (y_{1111}, \dots, y_{SJAT})'$ with Gaussian signal $\theta = (\theta_{1111}, \dots, \theta_{SJAT})'$ as given by (15). The signal θ is therefore a linear function of the state vector $\alpha = (\alpha_1, \dots, \alpha_T)'$ modelled by (14). The density function of the model is given by

$$p(y) = \int p(y, \theta) d\theta = \int p(y|\theta) p(\theta) d\theta, \quad (16)$$

where $p(y|\theta)$ is the Binomial density based on (8), that is $p(y|\theta) = \prod_t \prod_{s,j,a} p(y_{sjat} | \theta_{sjat})$. Furthermore, we have $p(\theta)$ as the Gaussian density. In case of non-Gaussian models, it is difficult to evaluate the integral (16) directly. Simulation techniques provide flexible and possibly efficient ways to numerically compute the likelihood function, that is $p(y)$ for a realisation of y . A standard simulation technique is importance sampling and is treated in the work of Ripley (1987), Kloek and Van Dijk (1978) and Geweke (1989). Implementations of importance sampling methods for state space time series models can be found in Shephard and Pitt (1997), Durbin and Koopman (1997) and in part II of Durbin and Koopman (2001). The basic idea behind the construction of the likelihood function is to sample from an importance density $g(\theta|y)$,

which is ideally as close as possible to density $p(\theta|y)$. The likelihood may then be written as

$$p(y) = \int p(y|\theta) \frac{p(\theta)}{g(\theta|y)} g(\theta|y) d\theta = E_g \left\{ p(y|\theta) \frac{p(\theta)}{g(\theta|y)} \right\}, \quad (17)$$

where E_g denotes expectation with respect to importance density $g(\theta|y)$. A Gaussian importance density is convenient since it is relatively straightforward to sample from. A linear Gaussian model that is close to the non-Gaussian model may be obtained by equalising the first and second derivatives of $\log p(y|\theta)$ and $\log g(y|\theta)$ with respect to θ . Solving the resulting (linear) equations will give $\hat{\theta} = E_g(\theta) = \int \theta g(\theta|y)$, see section 3.2 below.

Expression (17) can be simplified to the convenient formula

$$p(y) = g(y) E_g \left\{ \frac{p(y|\theta)}{g(y|\theta)} \right\}, \quad (18)$$

where $g(y)$ is the Gaussian likelihood function of the approximating model which can be obtained via the Kalman filter, see Durbin and Koopman (2001, §12.5) for further details.

An obvious estimator for the likelihood is $\hat{p}(y) = g(y)\bar{w}$, where

$$\bar{w} = \frac{1}{M} \sum_{i=1}^M w_i, \quad w_i = \frac{p(y|\theta^i)}{g(y|\theta^i)}, \quad (19)$$

and θ^i is a draw from importance density $g(\theta|y)$. In practice, the log of the likelihood function is used to manage the magnitude of density values. For this purpose, we evaluate $\log \hat{p}(y) = \log g(y) + \log \bar{w} +$ "bias correction" and maximize it to obtain maximum likelihood estimates of the parameters such as λ and β .

3.2 Constructing an importance density: the Binomial case,

To obtain an importance density we construct an approximating Gaussian model since generating samples for θ conditional on y is relatively straightforward when densities are normal. An approximating linear Gaussian model is obtained by linearization based on equalising the first and second derivatives of $\log p(y|\theta)$ and $\log g(y|\theta)$ (the latter is associated with approximating Gaussian model). The linearization takes place around a current guess of the mode of θ , say $\bar{\theta}^i$, and a new guess of the mode, say $\bar{\theta}^{i+1}$, is obtained from computing the conditional mean from the linear Gaussian model. Note that the mean and mode are the same in a Gaussian model. In more detail, we consider the Gaussian density $g(y|\theta)$ that corresponds to the linear Gaussian model

$$\tilde{y}_{s jat} = \theta_{s jat} + u_{s jat}, \quad u_{s jat} \sim \text{NID}(0, H_{s jat}), \quad (20)$$

for any permissible index values s, j, a and t , where $\tilde{y}_{s jat}$ and $H_{s jat}$ are obtained by the linearization conditions

$$\frac{\partial \log g(y|\theta)}{\partial \theta} = \frac{\partial \log p(y|\theta)}{\partial \theta}, \quad \frac{\partial^2 \log g(y|\theta)}{\partial \theta \partial \theta'} = \frac{\partial^2 \log p(y|\theta)}{\partial \theta \partial \theta'}.$$

We note that $p(\theta) \equiv g(\theta)$ and therefore the linearization also applies to the first and second derivatives of the densities $p(\theta|y)$ and $g(\theta|y)$ with respect to θ . It follows that \tilde{y}_{sjat} and H_{sjat} depend on θ and for the linearization they are functions of y_t and the current guess $\bar{\theta}^i$. In the case of the Binomial density we obtain

$$\tilde{y}_{sjat} = \theta_{sjat} + H_{sjat}y_{sjat} - 1 - \exp(\theta_{sjat}), \quad H_{sjat} = k_{sjat}^{-1}(1 + \exp(\theta_{sjat}))^2 \exp(-\theta_{sjat}), \quad (21)$$

for $\theta_{sjat} = \bar{\theta}_{sjat}^i$. Given values for \tilde{y}_{sjat} and H_{sjat} , the conditional mean of θ_{sjat} for the Gaussian model (20) can be computed by the Kalman filter and smoothing algorithm. The resulting estimate of θ is taken as the new guess of the mode, $\bar{\theta}^{i+1}$. The sequence of new estimates of θ converges to the mode of $p(\theta|y)$. For many different non-Gaussian log-densities, convergence to $\bar{\theta}_t$ is very quick and typically takes five to ten iterations.

The model (20) with \tilde{y}_{sjat} and H_{sjat} evaluated at the mode $\bar{\theta}_{sjat}$ is referred to as the ‘‘approximating’’ linear Gaussian model and the corresponding density $g(\theta|y)$ is taken as the importance density. By generating samples from $g(\theta|y)$ we effectively simulate θ ’s at the location of the mode of $p(\theta|y)$ and with a curvature implied by the approximating model. Drawing samples from $g(\theta|y)$ is based on model (20) and is relatively straightforward given the simulation smoothing algorithms of de Jong and Shephard (1995) and Durbin and Koopman (2002).

3.3 Signal extraction: estimation of factors

Given parameter estimates for λ and β , the unobserved signal may need to be estimated. In a similar way as we have obtained a Monte Carlo estimator of the likelihood function, we can obtain estimates of θ by evaluating the conditional mean

$$\bar{\theta} = \int \theta p(\theta|y) d\theta = p(y)^{-1} \int \theta p(\theta, y) d\theta = p(y)^{-1} \int \theta p(y|\theta)p(\theta) d\theta, \quad (22)$$

where density $p(y)$ is given by (18). The same motivation as for the Monte Carlo estimator $\hat{p}(y)$ applies such that the importance sampling estimator of $\bar{\theta}$ is given by

$$\hat{\bar{\theta}} = \frac{\sum_{i=1}^M \theta^i p(y|\theta^i) / g(y|\theta^i)}{\sum_{i=1}^M p(y|\theta^i) / g(y|\theta^i)}, \quad (23)$$

where θ^i is sampled from the importance density $g(\theta|y)$ for $i = 1, \dots, M$. The estimation of θ is referred to as signal extraction. When the signal consists of additive factors and covariates, they can be estimated separately using the same approach as presented here.

4 The data

4.1 Sample

The basic sample selected for the present study consists of 270 males who had been treated in the judicial institution Harreveld in the Netherlands, having been discharged between January

1989 to June 1996. Inclusion criteria were a minimum length of stay in the institution of 2 months, and a complete penitentiary treatment dossier. Treatment for sexual offending was an exclusion criterion. Juveniles may be treated in a judicial setting in the Netherlands for serious behavioral problems such as delinquency (under a criminal law measure) or because their home situation makes it impossible for them to remain at home (treatment is then imposed as a civil law measure). Formal consent for the study had been obtained from the Netherlands Ministry of Justice.

The sample cannot be considered representative for the population of males in juvenile institutions in the Netherlands: Boendermaker (1999), who compared samples from different juvenile institutions in the Netherlands, found juveniles in Harreveld over-representative for serious developmental problems. The group studied here may thus be considered a fairly serious group of juveniles. Descriptive analysis on psychological and background variables of this sample in an earlier study supports this (Van der Geest, Bijleveld, and Wijkman, 2005).

4.2 Psychological and background data

From penitentiary treatment files we extracted a number of items. These items include the dates of admission and discharge, (the type of) offences committed before admission, family background and personality characteristics. Items about family background contain, among other things, information about the criminal record of relatives and family unemployment. Intelligence scores (mostly obtained through the Wechsler Intelligence Scale for Children-Revised for the Netherlands (WISC-RN) or Raven Progressive Matrices) were turned into categories according to prevailing norm values. The youths' development of conscience was determined on the basis of data from the file and any pre-trial reports, mainly on the basis of clinician's statements.

4.3 Conviction data

Post-release offending information in this study is based on convictions in official data. We used Judicial Documentation (JD) abstracts of the Netherlands Ministry of Justice, that contain information on every case that is registered at the Public Prosecutor's Office and the ensuing verdict. Abstracts also contain information on date of commission and the type of offence. They span juvenile and adult offending, i.e. from age 12 up. Post release we were able to collect data on recidivism over an average period of 12.5 years, including shorter follow up times for 5 boys who had died young. On average, respondents were followed up onto age 30.

Using these data, for each respondent we coded per month after release whether he categorized as an offender. If so, we distinguished three types of offending: (0) No offending, (1) Non-serious offending, (2) Serious non-violent offending, (3) Violent offending. Using a standard classification for offenses in the Netherlands ("standaardclassificatie misdrijven" CBS, 2000),

we defined one or more of the following offenses as non-serious offending: vandalism, traffic offenses and all offenses under a special section of the criminal code (e.g. environmental offenses). We defined one or more of the following offenses as serious offending: property offenses, drug offenses, weapons violations, firearms regulations and arson. Violent offenses (including sexual offenses) are defined as the third, most serious, type of offending.

In case of multiple offenses per month, a score is based only on the most serious offense, irrespective to the number of offenses committed that month. As long as post-release conviction data is available, respondents score (0) in months where no offense was registered. Variables remain empty from the moment a respondent's conviction data are no longer available. Maximum observation time is almost 16 years. Most juveniles have a shorter observation time. Depending on the date of discharge, these juveniles have missing observations in the period 1989 to 1996. This results in a dataset of approximately 190 variables that represent a respondent's criminal career.

5 Empirical Results

In the following we present an example of the application of the non-Gaussian dynamic factor model to conviction data for 206 juveniles from the correctional facility of Harreveld. The findings we present on the basis of this model and data set should be viewed as illustrative and not as pertinent conclusions regarding the subject matter. There are many important background variables with strong interrelationships that we do not consider in the illustration below. Further, many categorical variables contain only a small number of individuals in important categories. We therefore present these results with appropriate reservation.

We first analyze the data presented above in Figure 1 with a plot of our proportions data for one type of crime, y_{2jat}/k_{2jat} , against time τ_t . The exact dates of the crimes are unknown to us, so the dates on the horizontal axis of Figure 1 are only indicative. These data concern the intensity of category 2 crimes, serious non-violent offending, per subgroup of juveniles as reported from the conviction data. The construction of the dependent variable only requires data on y_{2ajt} and k_{2ajt} . Category 2 crimes show the most interesting dynamics. They illustrate the main value added of our approach. Below, we also present results for crimes in the first and third category. In all cases we have left out the first sixteen and the last sixteen months of conviction data to reduce the influence of artificial boundary effects in the trend estimation. Our panel is very unbalanced, especially in the beginning of our sample and at the very end. We start with only 13 individuals in the first year. In the beginning of the sample we have many missing values as k_{2ajt} is often equal to zero, due to the later dates of discharge for most individuals. The maximum number of individuals is reached after 7 years and remains nearly constant until the end of our sample. The results remain qualitatively similar if we use a slightly longer time-series sample.

We examine the possibly time-varying heterogeneity due to two factors. First, the level of consciousness measured in the institution and second, the initial number of reported crimes before entering the institution. Data for these factors are available for a maximum of 206 individuals and were found to be significant determinants of crime in earlier studies of this data set. We split the sample in 3 times 4 subgroups as indicated in Table 1. In the context of model (2) we take $S = 1$, $J = 3$ and $A = 4$. The time series sample size is $T = 165$, and $\max_t \sum_j \sum_a k_{j,a,t} = 206$. We tried to incorporate deterministic age effects in the model by adding an age category variable, but the sample turned out to be too small to measure significant effects in this set-up.

For the dynamic component of the model we compare the continuous-time trend specification (4) for $f(\tau_t)$ with the discrete-time model (5) and an IID specification for f_t . Estimation of a discrete time AR(1) specification for $f(\tau_t)$ following Koopman and Lucas (2007) did not lead to satisfactory results, as the estimate of the autoregressive coefficient converged to 1. Instead of the model with one common component we also consider S and J independent trends below. All state space calculations were carried out using `SsfPack`, version 3 by Koopman, Shephard, and Doornik (1999), see also Durbin and Koopman (2001, Ch. 14). The data processing, likelihood optimisation and graphical analysis were done using the matrix programming language `Ox` of Doornik (1999).

Table 2 presents the estimation results for our preferred scalar common trends models for the binomial time series data for crime category 2, where $f(\tau)$ follows the continuous time smooth trend specification of (4), and where f_t follows the random walk specification of (5), respectively, which are assumed to be identical for each subcategory. We first discuss the estimates for the fixed effects $\lambda_{j,a}$. These estimates and their standard errors are very similar for the continuous time smooth trend model and the discrete time trend model. The initial number of crimes is a significant determinant. The juveniles with a low number of pre-stay crimes compared to the reference category of three or more crimes, continue to commit fewer crimes after leaving. It may seem surprising that boys with one pre-stay crime turn out to have a higher post-stay crime level than boys with two pre-stay crimes, but one has to realise that both of these categories contain only 18 cases. Here we used the original categorisation in the data, but there might be scope for additional pooling. Turning to the next explanatory variable we note that the level of consciousness has a negative effect: the individuals with low and lacunar conscience commit more crimes than the individuals with a sufficient conscience and individuals with low conscience turn out to be the most criminal. The estimates of the fixed effects seem sensible. Moreover, they are similar across different specifications for the dynamic part of the model, as will become apparent below. The dynamic part of the models in Table 2 contains only one parameter which measures the time variation in the scalar common trend, a time-varying nonstationary random effect. We can safely state that these models with a nonstationary random effect statistically outperform the model with an IID random effect

presented below in Table 7 as the log likelihood value differs by more than 40 points. The loglikelihood of the continuous-time smooth trend model is slightly lower than than for the discrete-time random walk specification, but this difference does not seem significant.

One has to bear in mind that formal testing for nonstationary components does not automatically follow from standard maximum likelihood (ML) theory, see, e.g. Nyblom and Harvey (2001) for (large T -) asymptotic theory on this topic in Gaussian State Space Models. Park and Phillips (2000) develop asymptotic theory in binary choice models for coefficient estimation with observable regressors following stochastic trends. In the latter situation, conventional ML based Wald tests apply in the usual way, although the convergence rate of the estimator is nonstandard and the limit distribution of the ML estimator is mixed normal. Even if the unobserved component is stationary one must take account of the possibility that one tests for a (variance) parameter value that it on the boundary of the maintained hypothesis, again leading to nonstandard asymptotic ML inference, see Andrews (2001).

The significance of the common trend factor is also confirmed by the smoothed estimates of $f(\tau)$ presented in Figure 2. As expected the overall crime level is not constant over calendar time, showing a significant decrease. The smooth continuous time trend starts to decline after a few years, whereas the discrete time trend starts to go down around the middle of our sample. Note that the scalings of the two trends and the corresponding β coefficients differ across the two specifications. The continuous time model implies a smoother function and tighter error bounds as it enforces a more gradual decline of the implicit local observation weights which define the trend estimate, see Durbin and Koopman (2001, section 4.6). Given the similar likelihood values for two trend models, either interpretation of the time variation might be valid.

The common downward trend might be a response to the overall aging of the individuals in the sample or the effect of more severe and more effective punishment or a response to decreasing returns on crime. Our data contains dates of birth of the individuals, but we were not able to effectively use this variable in our model. The individuals are roughly between 24 and 34 years old at the end of our sample. We do not have individual-specific or macro-data on other relevant time-varying explanatory variables for the crime careers.

Table 3 shows results for a vector common factor model, where β varies across the number of pre-stay offenses. There is some evidence for different rates of time variation if one applies a likelihood ratio test as the log likelihood is increased by nearly 8 points with the addition of 3 parameters. The Wald test on the equality of the β s does not indicate strong significance. The crime intensity of the juveniles with 2 pre-stay crimes seems to be more volatile. Table 4 presents the estimates for the discrete-time vector common factor model with different β s for different levels of consciousness. There is weak evidence that individuals with low and lacunar consciousness have a lower time variation in crime rates. The continuous-time specifications in Tables 3 and 4 show larger absolute t -values for the fixed effects and smaller absolute t -values

for the dynamic parameters, compared to the discrete time specification. Otherwise the vector common trend results are qualitatively similar across the two trend specifications.

Tables 5 and 6 present the independent trend alternatives of the models of Tables 3 and 4. The changes in the trend are now assumed to be independent across the different categories. As there are different trends for the different categories we remove the corresponding fixed effects from the model without a deterioration in model fit. The independent trend models do not outperform the vector common trend specifications. This is also illustrated by Figure 3, which shows the estimated independent dynamic factors corresponding to the model with independent trends for consciousness levels, compare Table 6. The estimated factors follow very similar trajectories. The decrease in crime seems to be most pronounced and the earliest for individuals with sufficient consciousness. This conclusion holds for both the smooth continuous time trend and the discrete time trend.

For comparison Table 7 shows the estimation results for the static scalar common factor model with an IID specification for f_t . The estimates for most λ 's are comparable to those of the previous models, but the model fit is significantly worse. If one believes the overall crime level to be stationary over time one can interpret the common λ in this model as the overall mean of the crime intensity, which is estimated with a high precision under this unrealistic assumption. The common λ does not have this interpretation in the nonstationary models that we have used. In sum, the simple scalar common trend models of Table 2 seem to do a decent job in describing the dynamics in crime rates for the second category. The model extensions do not lead to significant improvements in model fit or to new interpretations of the effects of the explanatory variables. The model can probably be simplified by pooling two initial crime history categories. The continuous-time and discrete-time trends provide similar results.

To conclude the presentation of empirical results, we present estimates of the scalar common factor model for crime categories 1 and 3 in Tables 8 and 9 and in Figures 4 and 5. Table 8 shows the results for the non-serious offenses. Comparing with the category 2 crime results in Table 2 we see a similar important effect of the pre-stay number of crimes, but the consciousness variable does not play a significant role. The common factor of category 1 crimes in Figure 4 is trending upwards, in contrast to the decreasing category 2 trend in Figure 2. The explanatory power is also relatively small. This conclusion holds for both trend models. It therefore does not seem to make empirical sense to pool the time series data of the first two crime categories. Table 9 shows the results for the violent crimes of category 3. Note that individuals that were treated for sexual offending were excluded from our sample, a priori. This may explain the absence of a significant effect of initial crime history on violent crime convictions. The consciousness variable is only marginally significant if we single out the lowest category. The corresponding Figure 5 shows no trend whatsoever in violent crimes. Apparently, there is not enough information in our data to identify an interesting dynamic factor model for violent crime careers.

6 Conclusions

We have modeled a panel data set of crime careers of juveniles from a Dutch judicial juvenile Institution. The data are decomposed into a systematic and an individual-specific component, of which the systematic component reflects the general time-varying conditions including the criminological climate. Within a model-based analysis, we treat shared effects of each group with the same background variables, strongly non-Gaussian features of the individual time-series, unobserved common systematic conditions, changing recidivism probabilities over time, and missing observations. We adopt a non-Gaussian multivariate state space model that deals with all of these issues simultaneously. We compare specifications in continuous time and discrete time. The parameters of the model are estimated by Monte Carlo maximum likelihood methods.

This paper illustrates the methods by presenting new empirical findings on the time variation in the recidivism behavior of approximately 200 juveniles over a period of 13 years. It turns out that the juveniles with a low number of pre-stay crimes compared to the reference category of three or more crimes, commit fewer serious crimes after leaving. We also note that the level of consciousness has a significant negative effect. Individuals with low and lacunar conscience commit more serious crimes than individuals with a sufficient conscience, and individuals with low conscience turn out to be the most serious criminals.

As expected the overall level of serious crime is not constant over calendar time, showing a significant decrease starting around the middle of our sample. This might be a response to the overall aging of the individuals in the sample or the effect of more severe and more effective punishment or a response to decreasing returns on serious crime. This systematic time variation turns out to be important and significant for serious crime, less important for non-serious crime and unimportant for violent crime.

In future research we shall investigate other specifications for the common trends. The examination of the empirical behaviour of asymptotic tests for the presence and the rank of the common component is also an interesting topic for further research, as indicated by the articles cited above in Section 5.

References

- Andrews, D. W. K. (2001). Testing when a parameter is on the boundary of the maintained hypothesis. *Econometrica* 69, 683–734.
- Bartolucci, F., F. Pennoni, and B. Francis (2007). A latent Markov model for detecting patterns of criminal activity. *Journal of the Royal Statistical Society, series A* 170, 115–132.

- Becker, G. (1968). Crime and punishment: An economic approach. *Journal of Political Economy* 76, 169–217.
- Bergstrom, A. R. (1984). Continuous time stochastic models and issues of aggregation over time. In Z. Griliches and M. D. Intriligator (Eds.), *Handbook of Econometrics, Vol. 2.*, Amsterdam, The Netherlands, pp. 1145–1212. North-Holland.
- Bijleveld, C. and A. Mooijaart (2003). Latent Markov modeling of recidivism data. *Statistica Neerlandica* 57, 305–320.
- Brame, R., E. Mulvey, and A. R. Piquero (2001). On the development of different kinds of criminal activity. *Sociological Methods and Research* 29, 319–341.
- Corman, H. and H. N. Mocan (2000). A time-series analysis of crime, deterrence, and drug abuse in New York City. *American Economic Review*, 90, 584–604.
- Cornwell, C. and W. Trumbull (1994). Estimating the economic model of crime with panel data. *Review of Economics and Statistics* 76, 360–366.
- de Jong, P. and N. Shephard (1995). The simulation smoother for time series models. *Biometrika* 82, 339–50.
- Doornik, J. A. (1999). *Object-Oriented Matrix Programming using Ox* (3rd ed.). London: Timberlake Consultants Press.
- D’Unger, A., K. Lund, P. McCall, and D. Nagin (1998). How many latent classes of Delinquent/Criminal careers? Results from mixed Poisson regression analyses. *American Journal of Sociology* 103, 1593–1630.
- Durbin, J. and S. J. Koopman (1997). Monte Carlo maximum likelihood estimation of non-Gaussian state space model. *Biometrika* 84, 669–84.
- Durbin, J. and S. J. Koopman (2001). *Time Series Analysis by State Space Methods*. Oxford: Oxford University Press.
- Durbin, J. and S. J. Koopman (2002). A simple and efficient simulation smoother for state space time series models. *Biometrika* 89, 603–616.
- Geweke, J. (1989). Bayesian inference in econometric models using Monte Carlo integration. *Econometrica* 57, 1317–39.
- Gupton, G., C. Finger, and M. Bhatia (1997). CreditMetrics - technical document, first edition. Technical report, RiskMetrics, www.riskmetrics.com.
- Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.
- Kloek, T. and H. K. Van Dijk (1978). Bayesian estimates of equation system parameters: an application of integration by monte carlo. *Econometrica* 46, 1–20.

- Koopman, S. J. and A. Lucas (2007). A non-Gaussian panel time series model for estimating and decomposing default risk. Technical report, Tinbergen Institute, Amsterdam, The Netherlands.
- Koopman, S. J., N. Shephard, and J. A. Doornik (1999). Statistical algorithms for models in state space using SsfPack 2.2. *The Econometrics Journal* 2, 107–160, www.ssfpack.com.
- Levitt, S. (1997). Using electoral cycles in police hiring to estimate the effect of police on crime. *American Economic Review* 87, 270–290.
- Nagin, D., D. Farrington, and T. Moffit (1995). Life-course trajectories of different types of offenders. *Criminology* 33, 111–139.
- Nagin, D. and K. Land (1993). Age, criminal careers and population heterogeneity: Specification and estimation of a nonparametric mixed-Poisson model. *Criminology* 31, 327–362.
- Nyblom, J. and A. C. Harvey (2001). Testing against smooth stochastic trends. *Journal of Applied Econometrics* 16, 415–429.
- Oud, J. L. and A. R. G. Jansen (2000). Continuous time state space modeling of panel data by means of SEM. *Psychometrika* 65, 199–215.
- Park, J. Y. and P. C. B. Phillips (2000). Nonstationary binary choice. *Econometrica* 68, 1249–1280.
- Piquero, A., R. Blumstein, R. Brame, R. Haapanen, E. Mulvey, and D. Nagin (2001). Assessing the effect of exposure time and incapacitation on longitudinal trajectories of criminal offending. *Journal of Adolescent Research* 16, 54–74.
- Ripley, B. D. (1987). *Stochastic Simulation*. New York: Wiley.
- Roeder, K., K. Lynch, and D. Nagin (1999). Modeling uncertainty in latent class membership: A case study in criminology. *Journal of the American Statistical Association* 94, 766–776.
- Schmidt, P. and A. D. Witte (1989). Predicting criminal recidivism using split population survival time models. *Journal of Econometrics* 40, 141–159.
- Shephard, N. and M. K. Pitt (1997). Likelihood analysis of non-Gaussian measurement time series. *Biometrika* 84, 653–67.
- Singer, H. (1998). Continuous panel models with time dependent parameters. *The Journal of Mathematical Sociology* 23, 77–98.
- Stander, J., D. Farrington, G. Hill, and P. Altham (1989). Markov Chain analysis and specialization in criminal careers. *British Journal of Criminology* 29, 317–335.
- Van der Geest, V. and C. Bijleveld (2006). Personal, background and treatment characteristics associated with offending after residential treatment: A 13-year follow up in

adolescent males. Technical report, Netherlands Institute for the Study of Crime and Law Enforcement (NSCR), Leiden, The Netherlands.

Van der Geest, V., C. Bijleveld, and M. Wijkman (2005). *Delinquentie na behandeling*. Technical report, Netherlands Institute for the Study of Crime and Law Enforcement (NSCR), Leiden, The Netherlands.

White, H., M. Bates, and S. Buyske (2001). Adolescence-limited versus persistent delinquency: Extending Moffitt's hypothesis into adulthood. *Journal of Abnormal Psychology* 110, 600–609.

Table 1: Frequency table background variables

	consciousness			
	low	lacunar	sufficient	
initial number				
of offenses				
0	6	16	29	51
1	2	9	7	18
2	2	10	6	18
3 or more	29	60	30	119
	39	95	72	206

Table 2: Explaining Crime Category 2 with Scalar Common Trend models

		Continuous Time Model (4)			Discrete Time Model (5)		
		coefficient	s.e.	<i>t</i> -value	coefficient	s.e.	<i>t</i> -value
λ	common	-3.01	0.120	-25.1	-3.26	0.241	-13.5
	0	-0.851	0.108	-7.86	-0.849	0.110	-7.72
	1	-0.142	0.128	-1.11	-0.141	0.134	-1.06
	2	-1.17	0.199	-5.88	-1.17	0.199	-5.88
	low	0.481	0.104	4.63	0.480	0.103	4.64
	lacunar	0.357	0.0914	3.90	0.359	0.0900	3.99
β	common	0.0873	0.0457	1.91	0.0593	0.0192	3.10
	log likelihood	-5846.07			-5845.95		

Note: β for Continuous time model multiplied by 100

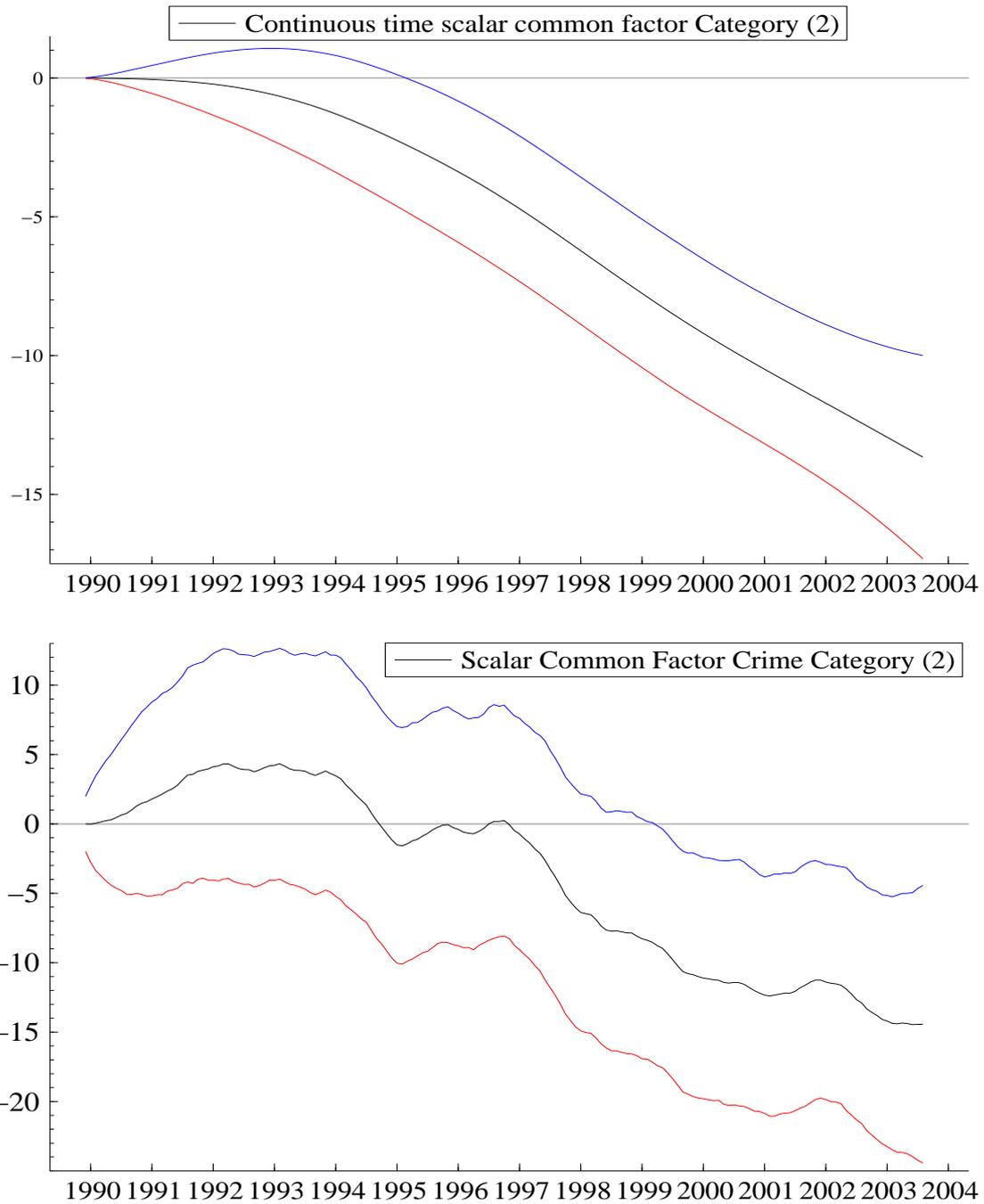


Figure 2: Smoothed estimates $\hat{f}(\tau_t) \pm 2$ s.e. for the Scalar Common Trend models explaining category 2 crime. $\hat{f}(\tau_t)$ multiplied by 100 for Continuous time model. Compare Table 2.

Table 3: Vector Common Trend models for Crime Category 2 with initial crime dependence

		Continuous Time Model (4)			Discrete Time Model (5)		
		coefficient	s.e.	<i>t</i> -value	coefficient	s.e.	<i>t</i> -value
λ	common	-3.16	0.111	-28.5	-3.36	0.208	-16.1
	0	-0.613	0.124	-4.94	-0.812	0.211	-3.85
	1	-0.178	0.127	-1.40	-0.158	0.134	-1.18
	2	-0.597	0.267	-2.24	-1.19	0.561	-2.12
	low	0.482	0.103	4.67	0.478	0.104	4.60
	lacunar	0.353	0.0910	3.88	0.355	0.0903	3.93
β	common	0.0925	0.0529	1.75	0.0527	0.0179	2.95
	0	0.0857	0.0617	1.39	0.0482	0.0258	1.87
	1	-0.00836	0.0400	-0.209	-0.00479	0.0222	-0.215
	2	0.239	0.171	1.40	0.138	0.0655	2.10
Wald test equal β 's		2.2626832			5.4073002		
log likelihood		-5838.65			-5838.07		

Note: β for Continuous time model multiplied by 100

Table 4: Vector Common Trend models for Crime Category 2 with initial consciousness dependence

		Continuous Time Model (4)			Discrete Time Model (5)		
		coefficient	s.e.	<i>t</i> -value	coefficient	s.e.	<i>t</i> -value
λ	common	-2.85	0.152	-18.8	-3.17	0.342	-9.29
	0	-0.856	0.108	-7.91	-0.854	0.109	-7.80
	1	-0.133	0.127	-1.04	-0.132	0.134	-0.988
	2	-1.16	0.199	-5.86	-1.17	0.200	-5.82
	low	0.175	0.121	1.44	0.335	0.205	1.64
	lacunar	0.114	0.104	1.10	0.260	0.161	1.61
β	common	0.145	0.0858	1.69	0.0911	0.0314	2.90
	low	-0.0757	0.0540	-1.40	-0.0493	0.0238	-2.08
	lacunar	-0.0620	0.0455	-1.36	-0.0382	0.0194	-1.97
Wald test equal β 's		2.028			4.650		
log likelihood		-5841.59			-5841.26		

Note: β for Continuous time model multiplied by 100

Table 5: Explaining Crime Category 2 with Independent Trends per initial crime category

		Continuous Time Model			Discrete Time Model		
		coefficient	s.e.	<i>t</i> -value	coefficient	s.e.	<i>t</i> -value
λ	common	-3.29	0.113	-29.0	-3.40	0.190	-18.0
	low	0.514	0.105	4.91	0.486	0.103	4.72
	lacunar	0.381	0.0911	4.18	0.365	0.0888	4.11
β	0	0.185	0.114	1.63	0.103	0.0403	2.56
	1	0.0844	0.0698	1.21	0.0761	0.0498	1.53
	2	0.278	0.159	1.75	0.180	0.0767	2.34
	≥ 3	0.103	0.0887	1.16	0.0551	0.0190	2.89
log likelihood		-5847.68			-5847.42		

Note: β for Continuous time model multiplied by 100

Table 6: Explaining Crime Category 2 with Independent Trends per level of consciousness

		Continuous Time Model			Discrete Time Model		
		coefficient	s.e.	<i>t</i> -value	coefficient	s.e.	<i>t</i> -value
λ	common	-2.73	0.0858	-31.7	-2.89	0.171	-16.9
	0	-0.870	0.109	-7.99	-0.856	0.109	-7.88
	1	-0.134	0.131	-1.02	-0.130	0.133	-0.980
	2	-1.16	0.203	-5.72	-1.17	0.199	-5.88
β	low	0.0690	0.0434	1.59	0.0482	0.0238	2.03
	lacunar	0.0742	0.0405	1.83	0.0553	0.0220	2.52
	sufficient	0.132	0.0682	1.93	0.0824	0.0295	2.80
log likelihood		-5844.18			-5846.47		

Note: β for Continuous time model multiplied by 100

Table 7: Explaining Crime Category 2 with a scalar Common IID Factor

		coefficient	s.e.	<i>t</i> -value
λ	common	-3.53	0.0807	-43.7
	0	-0.853	0.113	-7.56
	1	-0.169	0.126	-1.34
	2	-1.17	0.189	-6.16
	low	0.468	0.103	4.56
	lacunar	0.340	0.0861	3.95
β	common	0.336	0.0555	6.06
log likelihood		-5886.01		

Note: β for Continuous time model multiplied by 100

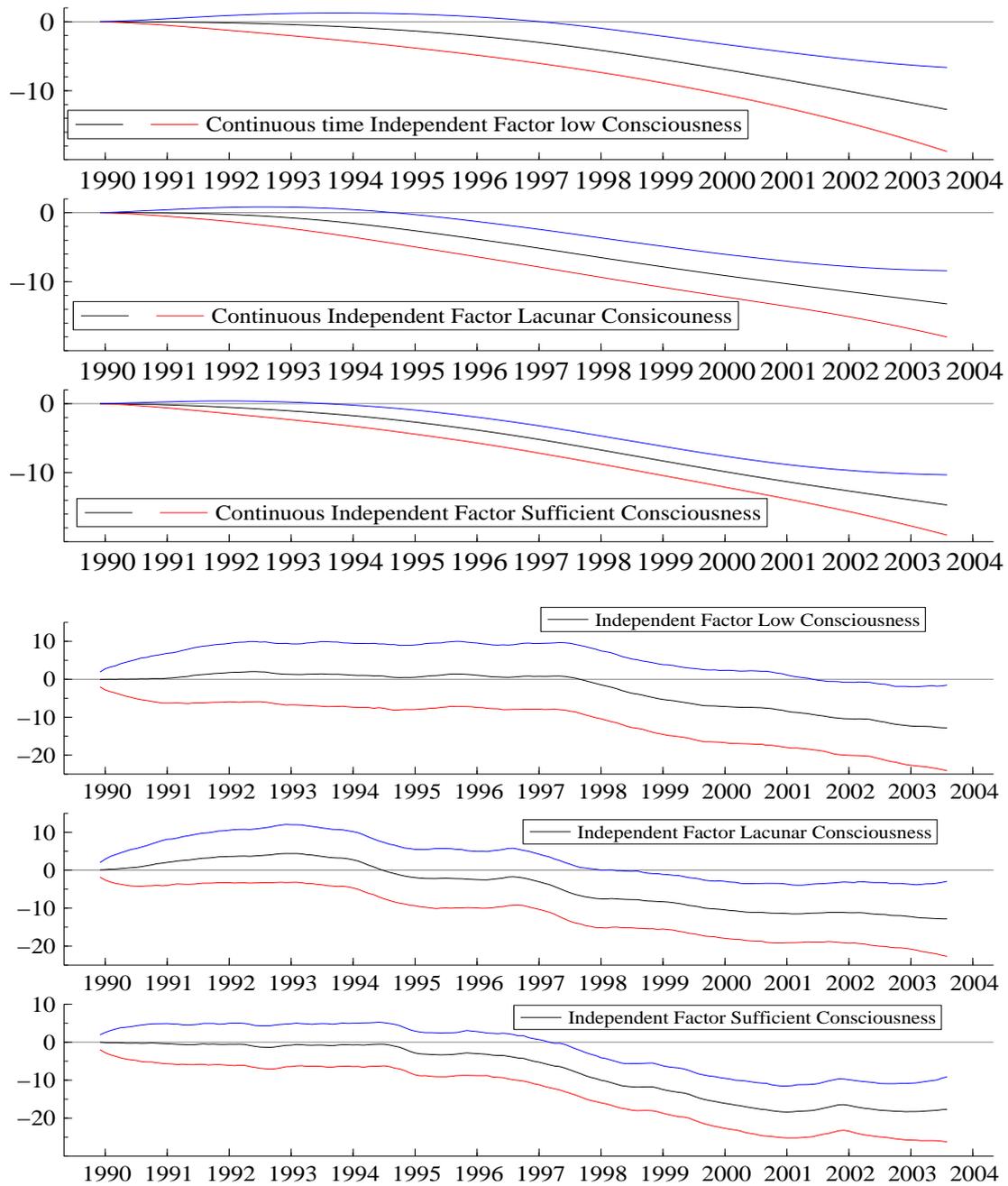


Figure 3: Smoothed estimates of $\hat{f}(\tau_t, j) \pm 2$ s.e. for the Independent Trends model explaining category 2 crime. $\hat{f}(\tau_t, j)$ multiplied by 100 for Continuous time model. Compare Table 6.

Table 8: Explaining Crime Category 1 with a Scalar Common Trend

		Continuous Time Model (4)			Discrete Time Model (5)		
		coefficient	s.e.	<i>t</i> -value	coefficient	s.e.	<i>t</i> -value
λ	common	-4.54	0.215	-21.1	-4.75	0.368	-12.9
	0	-0.535	0.122	-4.39	-0.535	0.121	-4.41
	1	-0.442	0.184	-2.40	-0.443	0.192	-2.31
	2	-1.11	0.244	-4.54	-1.11	0.246	-4.50
	low	0.177	0.122	1.45	0.178	0.120	1.47
	lacunar	-0.176	0.110	-1.61	-0.171	0.110	-1.55
β	common	0.178	0.0931	1.91	0.0755	0.0326	2.32
log likelihood		-4007.06			-4005.81		

Table 9: Explaining Crime Category 3 with Scalar Common Trend Models

		Continuous Time Model (4)			Discrete Time Model (5)		
		coefficient	s.e.	<i>t</i> -value	coefficient	s.e.	<i>t</i> -value
λ	common	-4.65	0.130	35.8	-4.66	0.132	-35.3
	0	-0.129	0.156	0.826	-0.129	0.159	-0.812
	1	-0.434	0.272	1.60	-0.438	0.289	-1.51
	2	-0.460	0.275	1.67	-0.463	0.280	-1.65
	low	0.370	0.171	2.16	0.369	0.171	2.16
	lacunar	-0.0113	0.155	0.0729	0.001	0.157	0.005
β	common	$2.8 * 10^{-6}$	0.022	$1.3 * 10^{-4}$	$-1.1 * 10^{-5}$	0.017	$-7 * 10^{-3}$
log likelihood		-2217.583			-2217.51		

Note: β for Continuous time model multiplied by 100

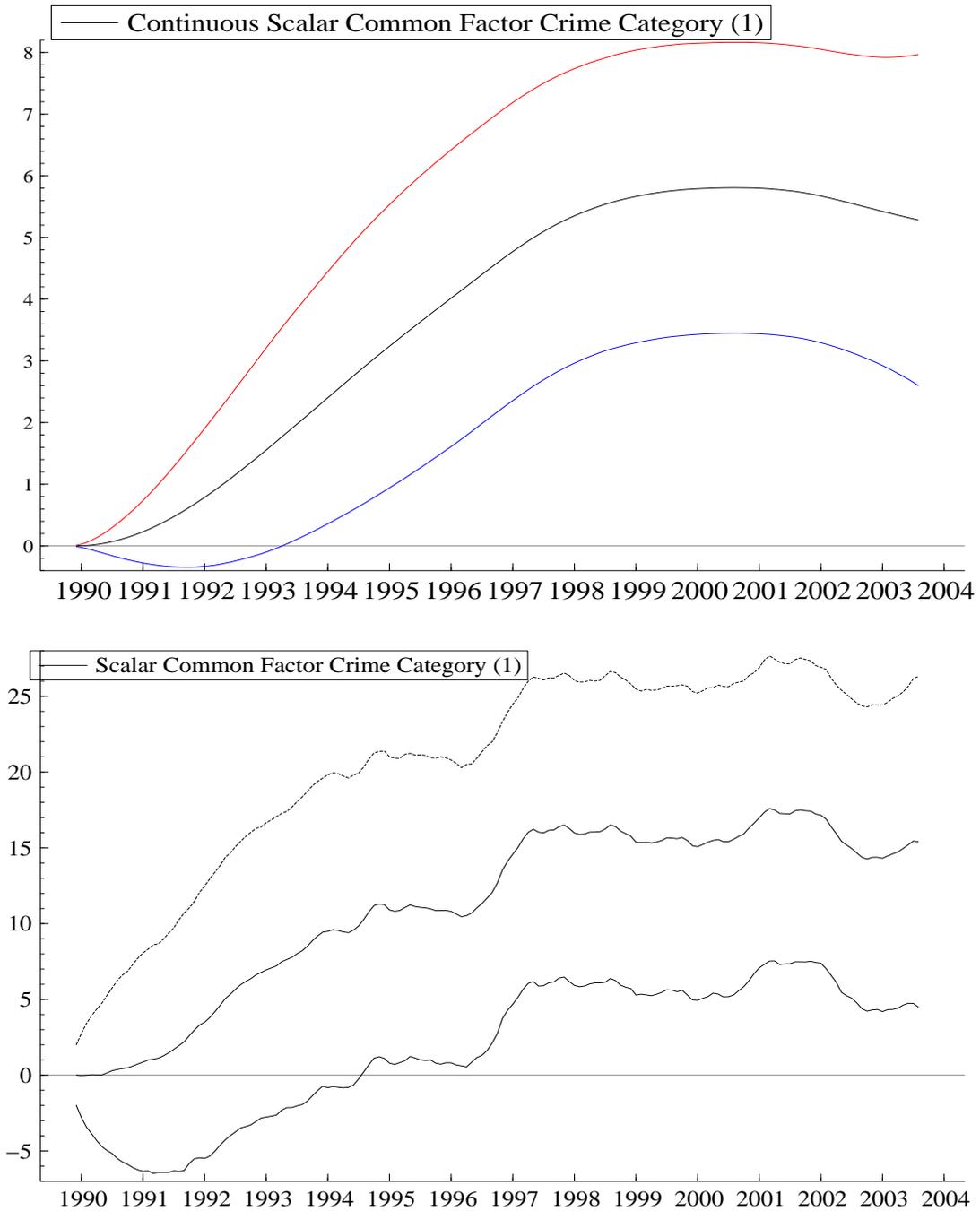


Figure 4: Smoothed estimates $\hat{f}(\tau_t) \pm 2$ s.e. for the Scalar Common Trend models explaining category 1 crime. $\hat{f}(\tau_t)$ multiplied by 100 for Continuous time model. Compare Table 8.

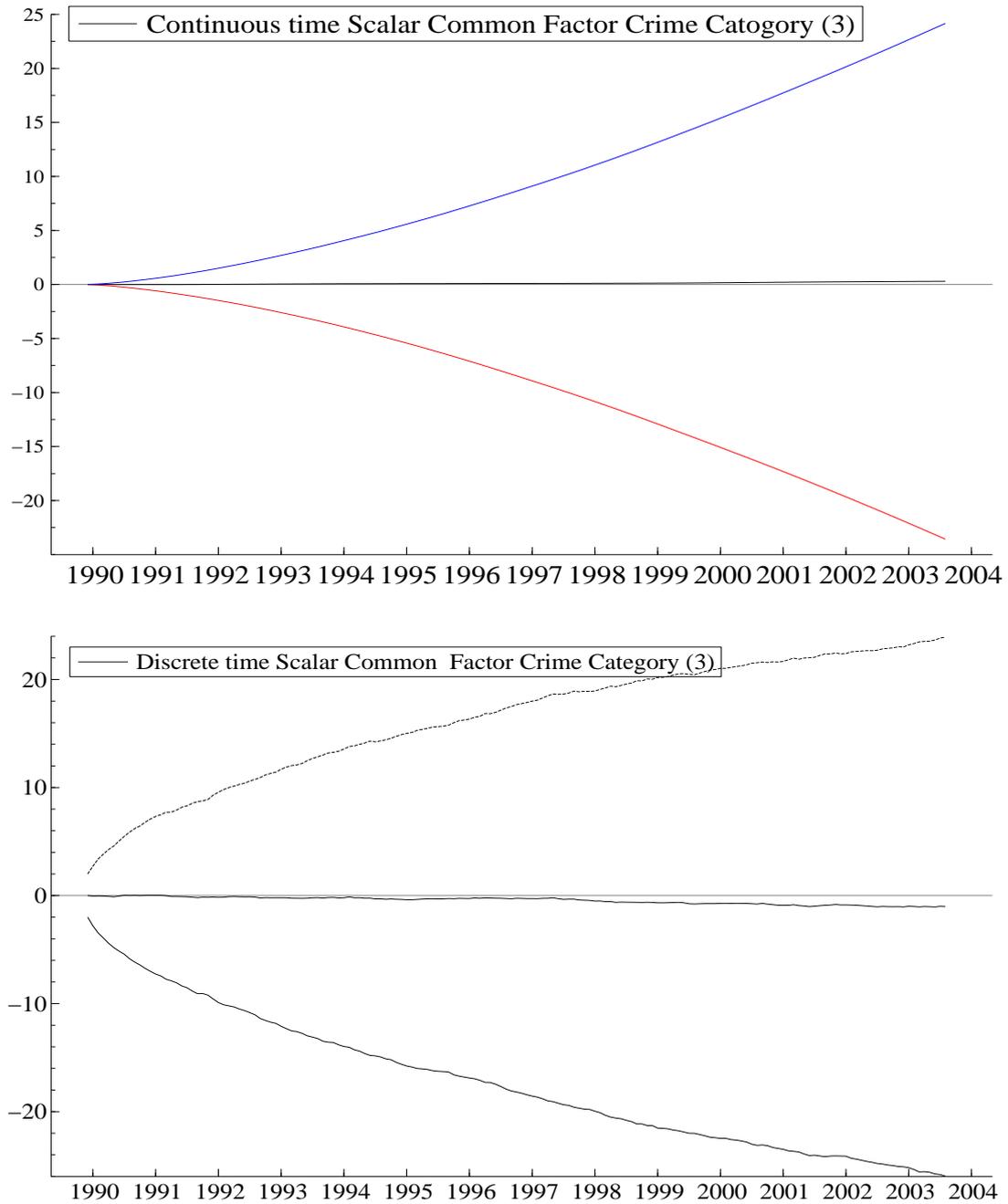


Figure 5: Smoothed estimates $\hat{f}(\tau_t) \pm 2$ s.e. for the Scalar Common Trend models explaining category 3 crime. $\hat{f}(\tau_t)$ multiplied by 100 for Continuous time model. Compare Table 9.