UPPSALA
UNIVERSITET

Modeling the Distribution of Financial Returns by Functional Data Analysis

Jonas Andersson

Paul Newbold

Division of Statistics

Research Report 2002:4

Uppsala University
Department of Information Science
P.O. Box 513
SE-751 20 Uppsala, Sweden

ISSN 1403-7572

Modeling the Distribution of Financial Returns by Functional Data Analysis

Jonas Andersson*and Paul Newbold†
Uppsala University and University of Nottingham

Abstract

In this paper, we use functional data analysis to model a time varying unconditional distribution of financial intraday returns. This is in the spirit of the recent development of realized volatility modeling (e.g. Andersen et al, 2001), where one of the moments of this uncondional distribution, the realized volatility, is assumed to change smoothly over time. In the approach used in this paper, we instead assume that the entire distribution function changes smoothly over time. This enables us to study auto- and cross dependencies of different parts of the unconditional distribution with no model assumptions but the smoothness of the distribution function. We develop a simulation based procedure for statistical inference of the model. Finally, we apply the method to the Swiss Franc-US Dollar exchange rate 1985-1991.

1 Introduction

In modeling financial returns, the approach usually taken is to assume strict stationarity and consider the moments of the distribution, conditional on previous observations, to be the quantities of interest. The most commonly used model of this kind is probably the generalized autoregressive conditional heteroskedasticity (GARCH) model, see Engle (1982) and Bollerslev (1986), where the conditional variance is modeled as a deterministic function of previous observations, but there are several other approaches, such as autoregressive stochastic volatility (SV) models (e.g., Clark, 1973; Taylor, 1986) and the hidden Markov model (HHM) (Hamilton, 1994). With the increased availability of so called tick data, i.e. stock prices or exchange rates observed for each single transaction, another kind of modeling, somewhat different in character, has become popular. This is the modeling of realized volatility (e.g. Andersen et al, 2001), which is defined as the sample variance of returns, calculated for each day. In a sense, when these sample variances have been calculated, the time series at hand is one that in most aspects can be modeled with standard conditional mean time series models. An important observation that can be

*Department of Information Science, Division of Statistics, Uppsala University, 751 20 Uppsala, Sweden. Helpful comments from Rolf Larsson and Johan Lyhagen are gratefully acknowledged. Also, many thanks to the School of Economics at the University of Nottingham for their hospitality during a stay on which the main part of this paper was written. Furthermore, the financial support from the Swedish Foundation for International Cooperation in Research and Higher Education for this stay is gratefully acknowledged.

†School of Economics, University of Nottingham , University Park, Nottingham, NG7 2RD, United Kingdom.

made about the realized volatility approach is that it is not to confused with the conditional variance. It is, as opposed to models of conditional variance, in fact calculated under the presumption that the unconditional variance for each day might change from day to day, i.e. the stationarity assumption of the return process has been abandoned.

A consequence of this way of modeling is that we might consider other properties of this time varying unconditional distribution, e.g. the kurtosis, quantiles, or indeed, the complete distribution function. The latter of these examples is the route taken in this paper. We will use the tools for functional data analysis (FDA) developed by Ramsay and Silverman (1997) in order to study the time dynamics of the daily unconditional probability distribution function. By contrast with GARCH and SV analysis, and in common with realized volatility analysis, we exploit intraday data. However, we study the entire distribution of returns rather than a single moment of it. Smoothness of evolution over time is allowed through an autoregressive specification.

In the next section, we will present the method of functional autoregression analysis. In order to fit the method to our particular application we will in Section 3 choose basis functions, appropriate for modeling inverse distribution functions. Section 4 shows how statistical inference for the model can be made by using the distribution of the empirical distribution function and in Section 5 we apply the method to intraday data on the Swiss Franc-US dollar exchange rate. Section 6 concludes the paper. A somewhat different application of functional data analysis - to the study of seasonal patterns in nondurable goods production - is given by Ramsay and Ramsey (2002).

2 The functional autoregressive model

In this section we will review the method developed in (Ramsay & Silverman, 1997) and taylor it to our application, i.e. a functional autoregression with one lag. We will also discuss two possible bases and outline how to estimate the model with more than one lag.

Consider the empirical inverse probability distribution function for time period t, $F_t(q)$, where $q \in (0,1)$, observed at times $t = 1, 2, ..., T$ and for the argument values $q_1, ..., q_N$. Let n_t denote the number of observations on which the function is based for time period t. These data, the quantiles, are stored in a $T \times N$ matrix

$$\mathbf{Y} = [F_t(q_j)]_{\substack{t=1,...,T \\ j=1,...,N}} \tag{1}$$

The reason for using the inverse distribution function rather than the distribution function or the probability density function is because of the technique used for making inference, explained in Section 4. Our goal is to study the time dynamics of this function, analogous to the situation when we want to estimate and interpret an autoregression of a time series $\{z_t\}_{t=1}^{T}$. We, therefore, state a functional autoregression model

$$\widehat{F}_t(p_0) = \alpha(p_0) + \int_0^1 F_{t-1}(p_1)\,\beta_1(p_1, p_0)\,dp_1 + ... + \int_0^1 F_{t-k}(p_1)\,\beta_k(p_1, p_0)\,dp_1. \tag{2}$$

Since we are mainly interested in the coefficient functions $\beta_i(p_1, p_0)$, $i = 1, ..., k$, we choose to work with the demeaned observations

$$F_t^*(p) = F_t(p) - \overline{F}(p) \tag{3}$$

where $\overline{F}(p)$ is the mean function

$$\overline{F}(p) = \frac{1}{T}\sum_{t=1}^{T} F_t(p). \tag{4}$$

The model is now rewritten as

$$\widehat{F_t^*}(p_0) = \int_0^1 F_{t-1}^*(p_1)\,\beta_1(p_1,p_0)\,dp_1 + ... + \int_0^1 F_{t-k}^*(p_1)\,\beta_k(p_1,p_0)\,dp_1 \tag{5}$$

and the intercept function $\alpha(p_0)$ can be calculated, after $\beta_1(p_1,p_0), ..., \beta_k(p_1,p_0)$ have been estimated, as

$$\widehat{\alpha}(p_0) = \overline{F}(p_0) - \int_0^1 \overline{F}(p_1)\left[\widehat{\beta}_1(p_1,p_0) + ... + \widehat{\beta}_k(p_1,p_0)\right] dp_1. \tag{6}$$

We will now review the procedure from Ramsay and Silverman's book on functional data analysis (Ramsay & Silverman, 1997), that we are going to use to estimate the parameter functions $\beta_1(p_1,p_0), ...,$ $\beta_k(p_1,p_0)$ in (5).

By letting the function be represented by some appropriate basis functions, $\phi_1, ..., \phi_J$, we will make it possible to estimate the functions at other values than $p_1, ..., p_N$. However, the main reason to employ this basis representation will become clear when the estimation procedure is being outlined. We write

$$F_t^*(p_0) = \sum_{j=1}^{J} c_{tj}\phi_j(p_0) \tag{7}$$

where J is the number of basis functions used. If we want our fitted function to match our data exactly at the observations we will have to use N basis functions. If we, on the other hand, are interested in smoothing the resulting prediction $\widehat{F_t^*}(p_0)$ we can use $J < N$. As pointed out by Ramsay and Silverman (1997), there are two reasons for using $J < N$. Firstly, as already mentioned, to smooth the result. Secondly, we want to avoid overfitting. In order to smooth the resulting prediction we reduce the number of basis functions for the dependent function and to avoid overfitting we reduce the number of basis functions for the regressors. There could thus be a reason to use a different number of basis functions for $F_t^*(p_0)$ than for the lags $F_{t-1}^*(p_1), F_{t-2}^*(p_2), .., F_{t-k}^*(p_k)$. We use K_l basis functions for the lags

$$F_{t-l}^*(p_l) = \sum_{j=1}^{K_l} c_{t-l,j}\phi_j(p_l) \tag{8}$$

and J basis functions for the dependent functional variable $F_t^*(p_0)$.

2.1 The case $k = 1$

First, we consider the case with $k = 1$. By putting the basis functions and their coefficients in matrices we can get a more compact notation for the basis representation

$$\mathbf{F}_0^*(p_0) = \mathbf{C}_0\boldsymbol{\phi}_0(p_0) \tag{9}$$

where

$$\mathbf{F}_0^*(p_0) = \left[F_t^*(p_0)\right]_{t=2,...,T} \tag{10}$$

$$\mathbf{C}_0 = \left[c_{tj}\right]_{\substack{t=2,...,T\\ j=1,...,J}} \tag{11}$$

and

$$\boldsymbol{\phi}_0(p_0) = \left[\phi_j(p_0)\right]_{j=1,\ldots,J}. \tag{12}$$

Analogously, we define

$$\mathbf{F}_1^*(p_0) = \mathbf{C}_1 \boldsymbol{\phi}_1(p_0) \tag{13}$$

where

$$\mathbf{F}_1^*(p_0) = \left[F_t^*(p_0)\right]_{t=1,\ldots,T-1} \tag{14}$$

$$\mathbf{C}_1 = \left[c_{tj}\right]_{\substack{t=1,\ldots,T-1\\ j=1,\ldots,K}} \tag{15}$$

and

$$\boldsymbol{\phi}_1(p_0) = \left[\phi_k(p_0)\right]_{k=1,\ldots,K}. \tag{16}$$

Further, we represent $\beta_1(p_1, p_0)$ in terms of the two systems of bases

$$\beta_1(p_1, p_0) = \sum_{j=1}^{J}\sum_{k=1}^{K} b_{jk}\phi_{1j}(p_1)\phi_{0k}(p_0) = \boldsymbol{\phi}_1(p_1)' \mathbf{B}\boldsymbol{\phi}_0(p_0) \tag{17}$$

where $\mathbf{B}$ is the matrix containing the coefficients of the expansion

$$\mathbf{B} = \left[b_{jk}\right]_{\substack{j=1,\ldots,J\\ k=1,\ldots,K}} \tag{18}$$

If the basis is not orthonormal we will also need the matrices

$$\mathbf{M}_0 = \left[\int_0^1 \phi_{0j}(p_1)\phi_{0k}(p_1)\,dp_1\right]_{\substack{j=1,\ldots,J\\ k=1,\ldots,J}} \tag{19}$$

and

$$\mathbf{M}_1 = \left[\int_0^1 \phi_{1j}(p_1)\phi_{1k}(p_1)\,dp_1\right]_{\substack{j=1,\ldots,K\\ k=1,\ldots,K}} \tag{20}$$

If we rewrite the predictions $\widehat{F_t^*}(x)$ by means of (5), we get

$$\begin{aligned}
\widehat{F_t^*}(p_0) &= \int_0^1 \left(\sum_{j=1}^{J} c_{1,t,j}\phi_{1j}(p_1)\sum_{l=1}^{J}\sum_{k=1}^{K} b_{lk}\phi_{1l}(p_1)\phi_{0k}(p_0)\right) dp_1 \\
&= \sum_{j=1}^{J}\sum_{l=1}^{J}\sum_{k=1}^{K} c_{1,t,j}\left(\int_0^1 \phi_{1j}(p_1)\phi_{1l}(p_1)\,dp_1\right) b_{lk}\phi_{0k}(p_0) \\
&= \sum_{j=1}^{J}\sum_{l=1}^{J}\sum_{k=1}^{K} c_{1,t,j}\mathbf{M}_{1jl} b_{lk}\phi_{0k}(p_0)
\end{aligned} \tag{21}$$

In matrix form, this can be written

$$\widehat{\mathbf{F}}_0^*(p_0) = \mathbf{C}_1\mathbf{M}_1\mathbf{B}\boldsymbol{\phi}_0(p_0) \tag{22}$$

The method used for estimation is to minimize the sum of the integrated squared residuals,

$$LMISE(\mathbf{B}) = \sum_{t=2}^{T}\int_0^1 \left(F_t^*(p_0) - \int_0^1 F_{t-1}^*(p_1)\beta_1(p_1, p_0)\,dp_1\right)^2 dp_0, \tag{23}$$

with respect to $\mathbf{B}$. By rewriting the summands of (23) as

$$\begin{aligned}
&\int_0^1 \left(F_t^*(p_0) - \int_0^1 F_{t-1}^*(p_1)\,\beta_1(p_1,p_0)\,dp_1\right)^2 dp_0 \\
=\;& \int_0^1 \left(\mathbf{c}_{0,t}'\boldsymbol{\phi}_0(p_0) - \mathbf{c}_{1,t}'\mathbf{M}_1\mathbf{B}\boldsymbol{\phi}_0(p_0)\right)\left(\boldsymbol{\phi}_0(p_0)'\mathbf{c}_{0,t} - \boldsymbol{\phi}_0(p_0)'\mathbf{B}\mathbf{M}_1\mathbf{c}_{1,t}\right)dp_0 \\
=\;& \int_0^1 \mathbf{c}_{0,t}'\boldsymbol{\phi}_0(p_0)\,\boldsymbol{\phi}_0(p_0)'\mathbf{c}_{0,t}dp_0 - \int_0^1 \mathbf{c}_{0,t}'\boldsymbol{\phi}_0(p_0)\,\boldsymbol{\phi}_0(p_0)'\mathbf{B}\mathbf{M}_1\mathbf{c}_{1,t}dp_0 \\
&- \int_0^1 \mathbf{c}_{1,t}'\mathbf{M}_1\mathbf{B}\boldsymbol{\phi}_0(p_0)\,\boldsymbol{\phi}_0(p_0)'\mathbf{c}_{0,t}dp_0 + \int_0^1 \mathbf{c}_{1,t}'\mathbf{M}_1\mathbf{B}\boldsymbol{\phi}_0(p_0)\,\boldsymbol{\phi}_0(p_0)'\mathbf{B}\mathbf{M}_1\mathbf{c}_{1,t}dp_0 \qquad (24)
\end{aligned}$$

where $\mathbf{c}_{0,t}$ and $\mathbf{c}_{1,t}$ are the rows of $\mathbf{C}_0$ and $\mathbf{C}_1$, respectively, which simplifies to

$$\begin{aligned}
&\int_0^1 \left(F_t^*(p_0) - \int_0^1 F_{t-1}^*(p_1)\,\beta_1(p_1,p_0)\,dp_1\right)^2 dp_0 \\
=\;& \left(\mathbf{c}_{0,t}' - \mathbf{c}_{1,t}'\mathbf{M}_1\mathbf{B}\right)\mathbf{M}_0\left(\mathbf{c}_{0,t}' - \mathbf{c}_{1,t}'\mathbf{M}_1\mathbf{B}\right) \qquad (25)
\end{aligned}$$

we obtain

$$LMISE(\mathbf{B}) = trace\left[(\mathbf{C}_1\mathbf{M}_1\mathbf{B} - \mathbf{C}_0)\,\mathbf{M}_0\,(\mathbf{C}_1\mathbf{M}_1\mathbf{B} - \mathbf{C}_0)'\right] \qquad (26)$$

which is known to be minimized by

$$\mathbf{B} = \mathbf{V}\boldsymbol{\Delta}^{-1}\mathbf{U}'\mathbf{C}_0 \qquad (27)$$

where $\mathbf{U}$, $\boldsymbol{\Delta}$ and $\mathbf{V}$ are determined by the singular value decomposition of $\mathbf{C}_1\mathbf{M}_1$,

$$\mathbf{C}_1\mathbf{M}_1 = \mathbf{U}\boldsymbol{\Delta}\mathbf{V}'. \qquad (28)$$

Now, when $\mathbf{B}$ is estimated, we can calculate $\beta_1(p_1,p_0)$ for any values of p_1 and p_0 we require. The function will give us a measure of how much the $100p_1\prime th$ quantile for day $t-1$ will affect the $100p_0\prime th$ quantile for day t.

2.2 The case for general k

We will now outline how to estimate the model for general k. The procedure is a direct extension of the case when $k=1$. We write

$$\widehat{F_t^*}(p_0) = \int_0^1 F_{t-1}^*(p_1)\,\beta_1(p_1,p_0)\,dp_1 + ... + \int_0^1 F_{t-k}^*(p_1)\,\beta_k(p_1,p_0)\,dp_1 \qquad (29)$$

and perform the same basis representation as in the previous section.

$$\widehat{\mathbf{F}}_0^*(p_0) = \mathbf{C}_1\mathbf{M}_1\mathbf{B}_1\boldsymbol{\phi}_0(p_0) + ... + \mathbf{C}_k\mathbf{M}_1\mathbf{B}_k\boldsymbol{\phi}_0(p_0) \qquad (30)$$

where $\mathbf{C}_i$ and $\mathbf{B}_i$ are the coefficient matrices of the basis expansion of the i'th lag and the i'th regression function, respectively. By defining the matrices

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_1 & \cdots & \mathbf{C}_k \end{bmatrix} \;(N \times Jk) \qquad (31)$$

and

$$\mathbf{B}^{(k)} = \begin{bmatrix} \mathbf{B}_1 \\ \vdots \\ \mathbf{B}_k \end{bmatrix} \;(Jk \times N) \qquad (32)$$

we can write (29) as

$$\widehat{\mathbf{F}}_0^*(p_0) = \mathbf{C}(\mathbf{M}_1 \otimes I_k)\,\mathbf{B}^{(k)}\boldsymbol{\phi}_0(p_0). \tag{33}$$

The matrices $\mathbf{B}_1$,...,$\mathbf{B}_k$ and consequently the regression functions $\beta_1(p_1,p_0)$,...,$\beta_k(p_1,p_0)$ can now be estimated in the same way as for the case $k=1$ by replacing $\mathbf{C}_1$ by $\mathbf{C}$ and $\mathbf{B}$ by $\mathbf{B}^{(k)}$.

3 Choice of basis

As mentioned above, we have to choose $K<N$ in order to avoid overfitting and $J<N$ to smooth the resulting predictions. The rationale for the latter dimension reduction is that we believe in a continuous distribution function and thus believe that an observed point on the function contains information on nearby points.

When the kind of basis is to be chosen one might argue, that basis functions that guarantee a non-negative first derivative of the prediction of $F_t^*(p)$, $dF_t^*(p)/dp$, is necessary. However, even though we acknowledge this point of view, we take the standpoint of Boneva et al (1971), who argue that the analytically tractable properties of Hilbert space methods outweigh the disadvantage of some small negative derivatives on the ground that if the absolute value of the derivative is very small we are not interested in these points of the distribution anyway.

3.1 Cubic spline basis

The first basis that we will consider is the cubic spline (CS) basis

$$\mathbf{CS}(p) = \begin{bmatrix} 1 & p & p^2 & p^3 & (p-\xi_1)_+^3 & \ldots & (p-\xi_m)_+^3 \end{bmatrix} \tag{34}$$

where $\xi_i = 1,...,m$ are the so called knots and $(x)_+ = 0$ if $x<0$ and x otherwise. These are chosen as the points $\xi_i = i/(m+1)$, $i=1,...,m$. We thus choose the basis functions to be

$$\begin{aligned}\boldsymbol{\phi}_0(p) &= \begin{bmatrix} 1 & p & p^2 & p^3 & (p-\xi_1)_+^3 & \ldots & \left(p-\xi_{J-4}\right)_+^3 \end{bmatrix}', & (35)\\ \boldsymbol{\phi}_1(p) &= \begin{bmatrix} 1 & p & p^2 & p^3 & (p-\xi_1)_+^3 & \ldots & \left(p-\xi_{K-4}\right)_+^3 \end{bmatrix}', & (36)\end{aligned}$$

This basis system has the property of giving the resulting function a continuous second order derivative, something that we feel is reasonable in our application to distribution functions.

The integrals needed to calculate $\mathbf{M}_0$ and $\mathbf{M}_1$

$$\mathbf{M}_0 = \left[\int_0^1 CS_n(p)\,CS_m(p)\,dp\right]_{n,m=0,\ldots,J-1} \tag{37}$$

and

$$\mathbf{M}_1 = \left[\int_0^1 CS_n(p)\,CS_m(p)\,dp\right]_{n,m=0,\ldots,K-1}, \tag{38}$$

where $CS_i(p)$ are the i'th element of $\mathbf{CS}(p)$, can be shown to be

$$\int_0^1 CS_n(p)\,CS_m(p)\,dp = \begin{cases} \frac{1}{n+m+1} \text{ if } 0\le n\le 3 \text{ and } 0\le m\le 3 \\ \frac{A(n,m)}{B(n)} \text{ if } 0\le n\le 3 \text{ and } m\ge 4 \\ \frac{A(m,n)}{B(m)} \text{ if } n\ge 4 \text{ and } 0\le m\le 3 \\ C(n,m) \text{ if } n\ge 4 \text{ and } m\ge 4 \end{cases} \tag{39}$$

(see the appendix for a derivation of these formulas) where

$$\begin{aligned} A(i,j) &= -6i^2 - 11i - 6 - i^3 + 21\xi_{j-3}i^2 + 42\xi_{j-3}i + 24\xi_{j-3} + \\ & 3\xi_{j-3}i^3 - 57\xi_{j-3}^2 i - 36\xi_{j-3}^2 - 24\xi_{j-3}^2 i^2 - 3\xi_{j-3}^2 i^3 + \\ & 26\xi_{j-3}^3 i + 24\xi_{j-3}^3 + 9\xi_{j-3}^3 i^2 + \xi_{j-3}^3 i^3 - 6\xi_{j-3}^{4+i}, \end{aligned} \quad (40)$$

$$B(i) = -(4+i)(3+i)(i+2)(i+1), \quad (41)$$

$$\xi_{\min} = \min(\xi_n, \xi_m), \quad (42)$$

$$\xi_{\max} = \max(\xi_n, \xi_m) \quad (43)$$

and

$$\begin{aligned} C(n,m) &= \frac{1}{140}(\xi_{\max}-1)^4(\xi_{\max}^3 - 7\xi_{\max}^2\xi_{\min} + 4\xi_{\max}^2 + 21\xi_{\max}\xi_{\min}^2 \\ & +10\xi_{\max} - 28\xi_{\max}\xi_{\min} + 20 - 70\xi_{\min} - 35\xi_{\min}^3 + 84\xi_{\min}^2). \end{aligned} \quad (44)$$

The coefficients, c_{tj}, $t = 1, ..., T$, and $j = 1, ..., J$ or K, are estimated by ordinary least squares.

The two tuning parameters that have to be subjectively chosen here are the number of basis functions and the positions of the knots. We do not intend to solve this problem here but will use a dataanalytic approach. We will try different numbers of basis functions and see if the results differ. Concerning the positions of the knots we will spread them evenly between zero and one.

3.2 Hermite polynomials

Another basis that we will evaluate is the one based on the Hermite polynomials. They are defined in terms of the standard normal distribution

$$H_n(p) = (-1)^n e^{p^2} D^n\left(e^{-p^2}\right), \quad (45)$$

where D^n is the $n'th$ derivative operator. The integral of the crossproducts that we need, can be shown to be

$$\int_0^1 H_n(p) H_m(p)\, dp = n!m! \sum_{j=0}^{\left[\frac{n}{2}\right]} \sum_{k=0}^{\left[\frac{m}{2}\right]} f_k(j) \quad (46)$$

(see the appendix) where

$$f_k(j) = \frac{(-1)^{j+k}}{2^{j+k} j!k!(n-2j)!(m-2k)!(n+m-2(j+k)+1)}. \quad (47)$$

The functions (47) can be calculated according to the recursion

$$f_{k+1}(j) = \frac{-(m-2k)(m-2k-1)(n+m-2(k+j))(n+m-2(k+j)-1)}{2(k+1)} f_k(j) \quad (48)$$

$$f_0(j) = \frac{(-1)^j}{2^j j!(n-2j)!m!(n+m-2j+1)} \quad (49)$$

Put in the notation used previously in this paper, this means that

$$\boldsymbol{\phi}_0(p) = [H_0(p), H_1(p), ..., H_J(p)]', \quad (50)$$

$$\boldsymbol{\phi}_1(p) = [H_0(p), H_1(p), ..., H_K(p)]', \quad (51)$$

$$\mathbf{M}_0 = \left[\sum_{j=0}^{\left[\frac{n}{2}\right]} \sum_{k=0}^{\left[\frac{m}{2}\right]} f_k(j)\right]_{n,m=0,1,...J-1} \quad (52)$$

and

$$\mathbf{M}_1 = \left[\sum_{j=0}^{\left[\frac{n}{2}\right]} \sum_{k=0}^{\left[\frac{m}{2}\right]} f_k(j) \right]_{n,m=0,1,\ldots K-1}. \tag{53}$$

The coefficients c_{tj}, $t = 1, ..., T$, and $j = 1, ..., J$ or K, are determined by ordinary least square estimation even though the orthogonality properties of the Hermite polynomials might be used to do this more efficiently. The reason we did not investigate this potential improvement here can be found in the application in Section 5 where the results for the spline basis and the Hermite basis were very similar.

4 Statistical inference

The method for statistical inference we propose is based on a simple fact. The uniform distribution of the empirical distribution function.

4.1 Two-standard error interval for $\beta_1(p_1, p_0)$

The method is a resampling method that is performed in the following way.

1. Estimate the inverse distribution functions $\widehat{F}_t(p)$ for each day $t = 1, ..., T$ by using the chosen basis.
2. Draw n_1 random numbers from a uniform distribution $p_{11}, p_{12}, ..., p_{1n_1} \sim U(0,1)$ where n_1 is the number of returns in day one.
3. Use the estimated inverse distribution function to calculate a resample $\mathbf{x}_1 = \left[\widehat{F}_t^*(p_{11}), ..., \widehat{F}_t^*(p_{1n_1})\right]'$ for day one.
4. Estimate the inverse distribution function $\widehat{F}_1^*(p)$ of the resample $\mathbf{x}_1$ by using the chosen basis.
5. Repeat step 2 - step 3 for $t = 2, ..., T$.
6. Calculate the first resample $\mathbf{Y}_1^* = \left[\widehat{F}_t^*(p_j)\right]_{\substack{t=1,...,T \\ j=1,...,N}}$.
7. Estimate the model, but replace $\mathbf{Y}$ with $\mathbf{Y}_1^*$. Save the result ${\xi_{\min}}_1^*$.
8. Repeat steps 3-7 N_{repl} times in order to get ${\xi_{\min}}_1^*, {\xi_{\min}}_2^*, ..., {\xi_{\min}}_{N_{repl}}^*$.

By using the estimated ${\xi_{\min}}^*$ matrices and (17) we can now calculate N_{repl} functions $\beta_{11}^*(p_1, p_0), ..,$ $\beta_{1N_{repl}}^*(p_1, p_0)$, all considered to be drawn from the same distribution of functions. By calculating the variance function of these we can form 2-standard error intervals around the estimated regression function.

$$\widehat{\beta}_1(p_1, p_0) \pm 2\sqrt{\widehat{var}\left(\widehat{\beta}_1(p_1, p_0)\right)} \tag{54}$$

where

$$\widehat{var}\left(\widehat{\beta}_1(p_1, p_0)\right) = \frac{1}{N_{repl}} \sum_{i=1}^{N_{repl}} \left(\beta_{11}^*(p_1, p_0) - \overline{\beta}_1^*(p_1, p_0)\right)^2 \tag{55}$$

and $\overline{\beta}_1^*(p_1, p_0)$ is the mean function of the resampled coefficient functions.

5 Application to exchange rates

5.1 The data

The data are the extended version of the tickwisely observed Swiss Franc - US Dollar data that were used in the Santa Fe forecasting competition 1991.[1] The sample period is the 20th of May 1985 to the 12th of April 1991. The mean of the time elapsed between transactions is a little more than two and a half minutes and we chose to extract the prices each 5 minutes and calculate the first difference of the log prices. We also remove observations before 8 am and after 17 pm in order to get the same number of quotes each day. The prices are calculated as the means of the bid and ask prices. Since we would estimate the unconditional distribution function day by day and thus for each day wanted a sample that could be considered to be taken from a stationary process, we removed the overnight returns. This being done, the $k * 5\%$ percentiles where calculated for $k = 1, ..., 19$ for each day. This means that we now have 19 time series of 1479 days of observations each. These percentiles are the ones that are later going to be considered as observed points of the functions that will be modeled in the functional data analysis. Because of the illustrative character of this application we will limit the analysis by just using one lag.

We first consider the first order autocorrelation coefficients of the quantiles. As can be seen from Figure 1 the first-order autocorrelation coefficient is smallest for the quantiles in the middle of the distribution. The general impression of the graph is that the strongest first-order autocorrelation is found in the tails.

Figure 1 about here

5.2 Estimation of the model

The point estimate of the regression function was made both with the cubic spline basis and the Hermite basis, see Figure 2. Since the results were very similar we choose to present, in detail, only the cubic spline results here. The choice of number of knots is obviously arbitrary so one can at best hope that the choice does not alter the interpretation of the result. We deal with this problem by doing the analysis with two choices, three and seven knots. These choices correspond to seven and eleven basis functions respectively. These numbers should be compared with the number of points used when calculating the function, which was 19.

If we first consider the three dimensional graphs in Figure 2 and Figure 3, illustrating the entire estimated regression function, we can see that it corresponds to Figure 1 on the diagonal, separately plotted in Figure 4. The diagonal of the regression function, i.e. the function value when the arguments p_1 and p_0 are equal, has peaks at the 25th and 75th percentiles. As opposed to this symmetry, the diagonal is increasing both for small and large values. The implication of this is that the further out in the right tail we move, the more first-order time dependency we find. This is as opposed to the left tail where the first-order time dependency decreases when we move further out in the tail. This result could also be observed in Figure 1.

[1] The data was downloaded from the webpage http://www-psych.stanford.edu/~andreas/Time-Series/Data/.

Figure 2, 3 and 4 about here

The function in Figure 4 is rather small if $p_0 \approx 0.5$ which, together with observation of figures 5 to 9 means that the function's dependency over time is mainly manifested in the tails of the distribution and that there is a negative dependency between the two tails. This corresponds to results usually obtained with conventional volatility models such as GARCH and SV models.

The regression function contains more information than Figure 1 since we also can see dependencies over time between different parts of the distribution. In the two dimensional plots in figures 5 to 9, p_1 has been held fixed. We show the graphs for the values $p_1 = 0.05$, 0.25, 0.5, 0.75 and 0.95 together with their 2-standard error bands.

Figures 5-9 about here

The general impression is, as expected, that the time dependency is weak in the middle of the distribution, see Figure 7. The notion of time dependency here means not only autocorrelation of a certain percentile but also dependency between different percentiles for consequtive days. That the time dependency is weak in the middle of the distribution thus means that the dependency between the 50th percentile for day $t-1$ and all percentiles for day t is weak.

The strong time dependency in the two tails is different in one aspect. While the strongest dependency in the left tail is found around the 25th percentile, the percentiles which are most dependent on the distribution function for time $t-1$ in the right tail are the ones around the 95th quantile.

The dependency is asymmetric in the sense that the time dependency is stronger in the left tail of the distribution than in the right. It is on the other hand symmetric for each p_1 in the sense that the value of the regression function for the p_1th percentile at time t and the p_0th percentile at time $t-1$ is about the same as with the $(1-p_0)$th percentile but with the opposite sign, i.e. $\beta(p_1, p_0) \approx -\beta(p_1, 1-p_0)$.

6 Conclusions

We have used the functional regression method of Ramsay and Silverman (1997) to estimate a functional autoregression of the daily inverse distribution function of the intraday changes of the Swiss Franc - US Dollar exchange rate. We have introduced and used a simulation based procedure to calculate 2-standard error intervals for the estimated regression function. By fully using the intraday data, compared to only daily data, information in these intraday data can be used to model the entire distribution function nonparametrically. We can thereby, with no distributional assumptions, except continuity of the distribution function, study the dynamics of this distribution. We have used the assumed continuity of the distribution function and regressed it on the corresponding function for past days by functional regression analysis. The method is to a large extent nonparametric with the exception of the use of the continuity assumption, something that might be considered as a model assumption.

This way of modeling the returns has significant differences from conventional approaches such as GARCH and stochastic volatility modeling. The most important difference is that in order to investigate a known and discover new so called stylized fact of the dynamic and distributional properties of financial returns we do not need to assume a parametric model describing the stylized fact we want to study. It has also one significant difference from pure descriptive approaches. It uses the smoothness of the distribution function. A method it reseambles to a certain extent is the modeling of realized volatility which, in fact, is one of the moments in this distribution.

The empirical results indicate that the strength of the day-to-day dependency in the data set we study over time is strongest in the 25th and 95th percentile. The 25th percentile for day t has its strongest dependency with the 25th and the 75th percentile for day $t-1$ while the 95th has its strongest dependency with the 5th and 95th percentile for day $t-1$. The dependecy is, as expected, very small in the middle of the distribution but also, which was less expected, far out in the left tail. This is clearly a richer set of conclusions than could follow from an analysis concentrating exclusively on variances.

References

Andersen, T.G., Bollerslev, T., Diebold, F.X. and Labys, P. (2001) The distribution of realized exchange rate volatility. *Journal of the American Statistical Association* , 96, 42-55.

Boneva, L.I., Kendall, D. and Stefanov, I. (1971) Spline transfromations: Three new diagnostic aids for the statistical data-analyst. *Journal of the Royal Statistical Society* , Ser B, 33, 1-70.

Bollerslev, T. (1986) Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics* 51, 307-327.

Clark, P.K. (1973) A subordinated stochastic process model with finite variance for speculative prices. *Econometrica* 41, 135-156.

Engle, R.F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of the United Kingdom inflation. *Econometrica* 50, 987-1007.

Falk, M. (1984) Relative defiency of kernel type estimators of quantiles. *Annals of Statistics* 23, 261-268.

Hamilton, J.D. and Susmel, R. (1994) Autoregressive Conditional Heteroskedasticity and Changes in Regime. *Journal of Econometrics* 64, 307-334.

Lange, K. (1999) *Numerical analysis for statisticians.* Springer, New York.

Taylor, S.J. (1986) *Modelling financial time series.* John Wiley, Chichester.

Ramsay, J.O. and Ramsey, J.B. (2002) Functional data analysis of the dynamics of the monthly index of nondurable goods production. *Journal of Econometrics* 107, 327-344.

Ramsay, J.O. and Silverman, B.W. (1997) *Functional data analysis.* Springer, New York.

Appendix

The matrices M_0 and M_1

Cubic spline basis

The J-dimensional spline basis can be decomposed in two parts

$$\boldsymbol{\phi}_{pol}(p) = \begin{pmatrix} 1 & p & p^2 & p^3 \end{pmatrix} \quad (56)$$

and

$$\boldsymbol{\phi}_{knot}(p) = \begin{pmatrix} (p-\xi_1)_+^3 & \cdots & (p-\xi_{J-4})_+^3 \end{pmatrix} \quad (57)$$

The integrated cross products of combinations of the elements in $\boldsymbol{\phi}_{pol}(p)$ can be calculated as

$$\int_0^1 p^{n+m} dp = \frac{1}{n+m+1} \quad (58)$$

for $n, m = 0, 1, 2$ and 3.

For combinations of one element in $\boldsymbol{\phi}_{pol}(p)$ and one in $\boldsymbol{\phi}_{knot}(p)$ the integrals become

$$\int_0^1 p^n \left(p-\xi_{m-3}\right)_+^3 dp = \int_{\xi_{m-3}}^1 p^n \left(p-\xi_{m-3}\right)^3 dp \quad (59)$$

which can be shown to be as the second row in equation (39). The third row in (39) is as the second but with n and m interchanged.

Finally, the cross product integrals for combinations of elements in $\boldsymbol{\phi}_{knot}(p)$ we write

$$\int_0^1 \left(p-\xi_{n-3}\right)_+^3 \left(p-\xi_{m-3}\right)_+^3 dp = \int\limits_{\max(\xi_{n-3},\xi_{m-3})}^{1} \left(p-\xi_{n-3}\right)^3 \left(p-\xi_{m-3}\right)^3 dp \quad (60)$$

which can be shown to be solved as the last row of (39).

Hermite polynomial

The Hermite polynomials

$$H_n(p) = (-1)^n e^{p^2} D^n \left(e^{-p^2}\right), \quad (61)$$

where D^n is the n'th derivative operator, can be calculated according to

$$H_n(p) = \sum_{j=0}^{\left[\frac{n}{2}\right]} \frac{n!(-1)^j p^{n-2j}}{2^j j!(n-2j)!} \quad (62)$$

(Lange, 1999). Consequently the cross product $H_n(p)H_m(p)$ can be written

$$H_n(p)H_m(p) = n!m!p^{n+m} \sum_{j=0}^{\left[\frac{n}{2}\right]} \sum_{k=0}^{\left[\frac{m}{2}\right]} \frac{(-1)^{j+k} p^{-2(j+k)}}{2^{j+k} j!k!(n-2j)!(m-2k)!}. \quad (63)$$

By integrating over (63), we obtain

$$\int_0^1 H_n(p)H_m(p)dp = n!m! \sum_{j=0}^{\left[\frac{n}{2}\right]} \sum_{k=0}^{\left[\frac{m}{2}\right]} \frac{(-1)^{j+k}}{2^{j+k} j!k!(n-2j)!(m-2k)!(n+m-2(j+k)+1)} \quad (64)$$

Since the values of the denominator in (64) can become large when a large number of basis functions is used, the following recursion could be useful. Write

$$f_k(j) = \frac{(-1)^{j+k}}{2^{j+k} j! k! (n-2j)! (m-2k)! (n+m-2(j+k)+1)}. \tag{65}$$

Then

$$\begin{aligned}
f_{k+1}(j) &= \frac{(-1)^{j+k+1}}{2^{j+k+1} j!(k+1)! (n-2j)! (m-2(k+1))! (n+m-2(j+k+1)+1)} \\
&= \frac{-1(m-2k)(m-2k-1)(n+m-2(k+j))(n+m-2(k+j)-1)}{2(k+1)} \\
&\quad \times \frac{(-1)^{j+k}}{2^{j+k} j! k! (n-2j)! (m-2k)! (n+m-2(j+k)+1)} \\
&= \frac{-(m-2k)(m-2k-1)(n+m-2(k+j))(n+m-2(k+j)-1)}{2(k+1)} f_k(j)
\end{aligned}$$

Figures

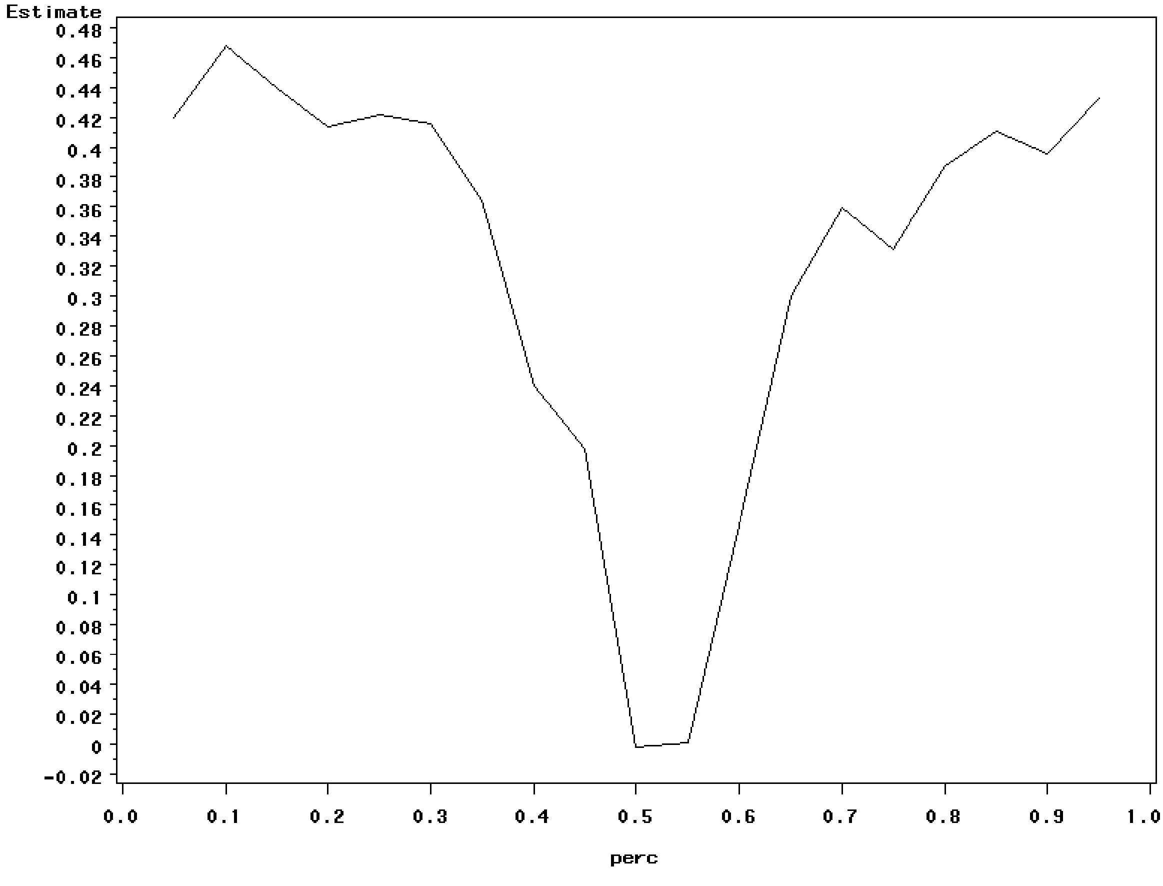

Figure 1. First-order autocorrelation coefficients for different quantiles. Returns calculated as first difference of log prices.

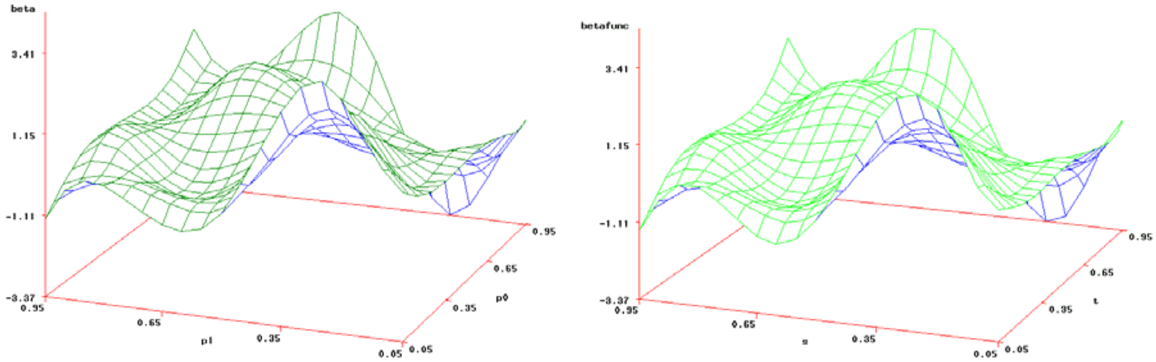

Figure 2. The estimated regression function when 7 basis functions were used. The bases used are from left to right the spline- and Hermite bases.

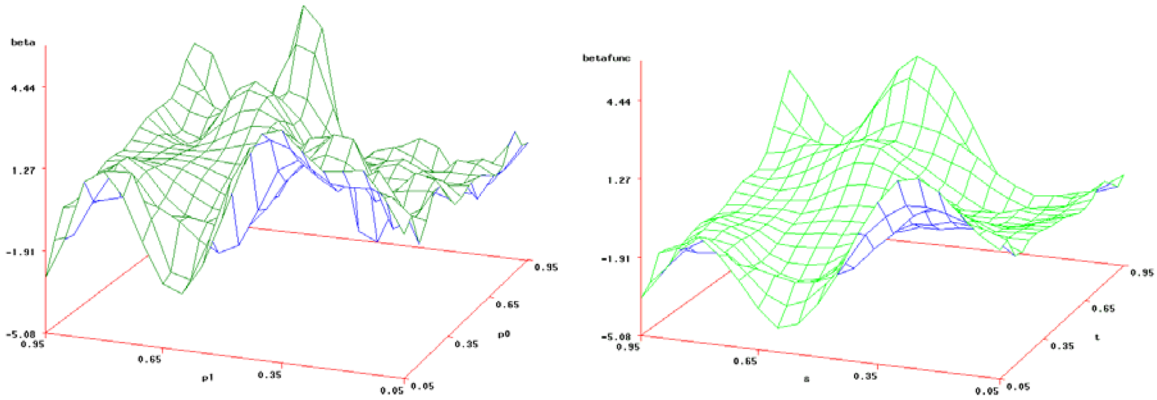

Figure 3. The regression function when 11 basis functions were used. The bases used are from left to right the spline- and Hermite bases.

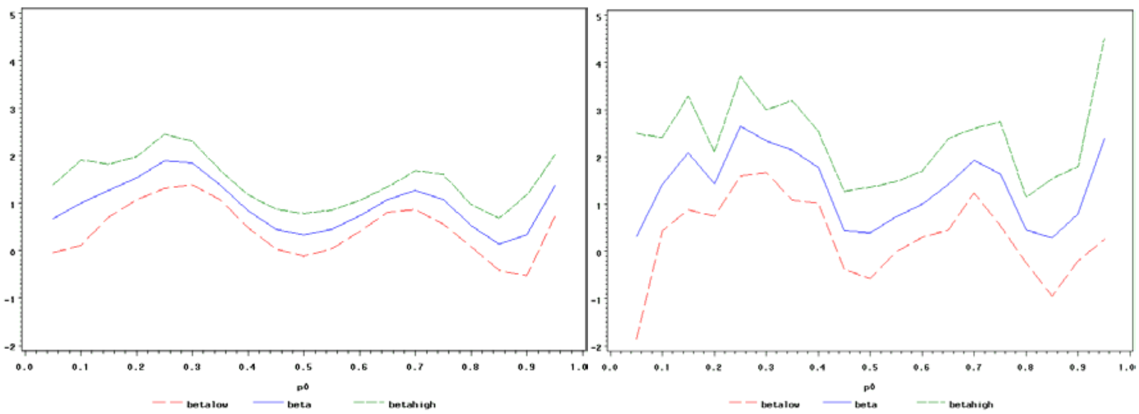

Figure 4. The diagonal of the regression function, i.e. the cross section of the function when $p_0 = p_1$, when 3 and 7 knots were used. The three lines represent the estimated regression function and their 2-standard error intervals.

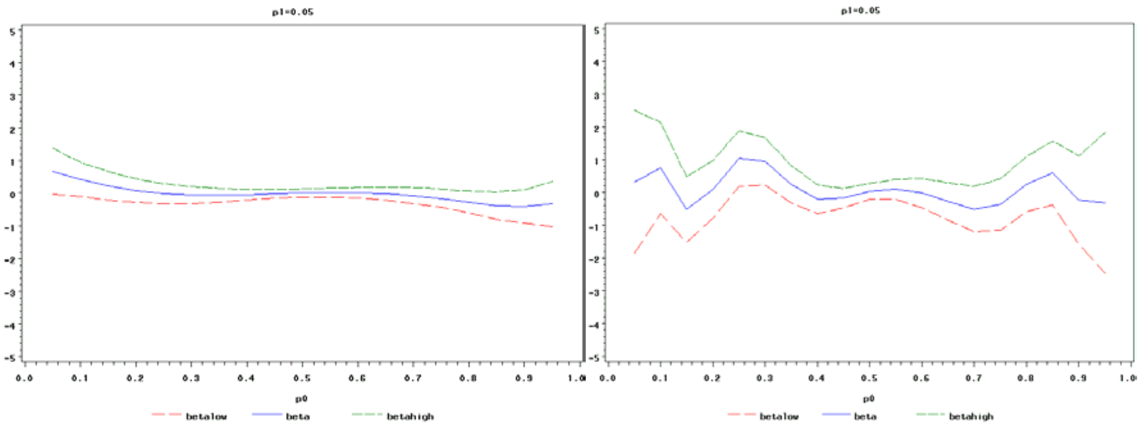

Figure 5. Marginal plot of the regression function on the percentile at time t, p_0 for the 5th percentile at time $t-1$. The number of knots used in the left graph was 3 and in the right 7. The three lines represent the estimated regression function and the 2-standard error intervals.

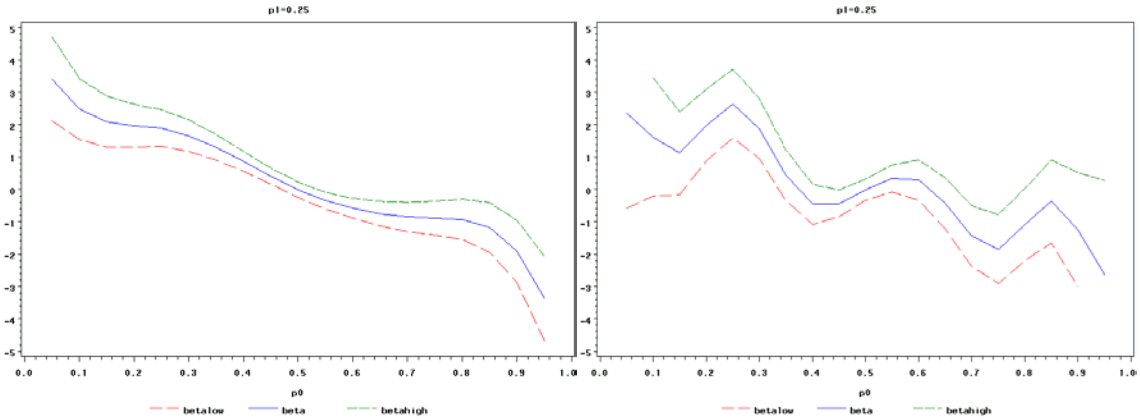

Figure 6. Marginal plot of the regression function on the percentile at time t, p_0 for the 25th percentile at time $t-1$. The number of knots used in the left graph was 3 and in the right 7. The three lines represent the estimated regression function and their 2-standard error intervals.

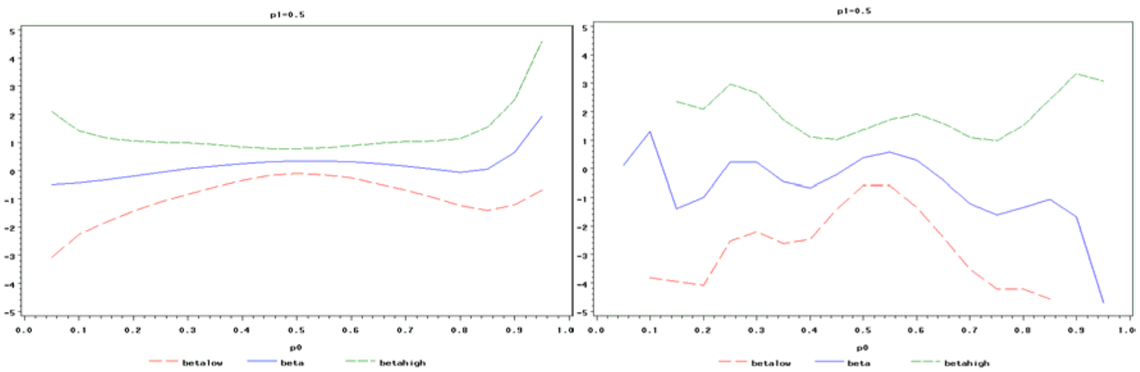

Figure 7. Marginal plot of the regression function on the percentile at time t, p_0 for the 50th percentile at time $t-1$. The number of knots used in the left graph was 3 and in the right 7. The three lines represent the estimated regression function and their 2-standard error intervals.

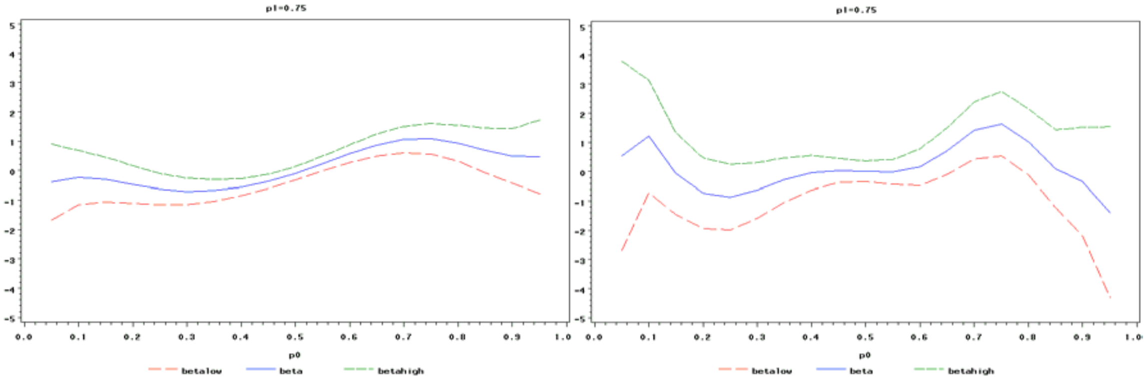

Figure 8. Marginal plot of the regression function on the percentile at time t, p_0 for the 75th percentile at time $t-1$. The number of knots used in the left graph was 3 and in the right 7. The three lines represent the estimated regression function and their 2-standard error intervals.

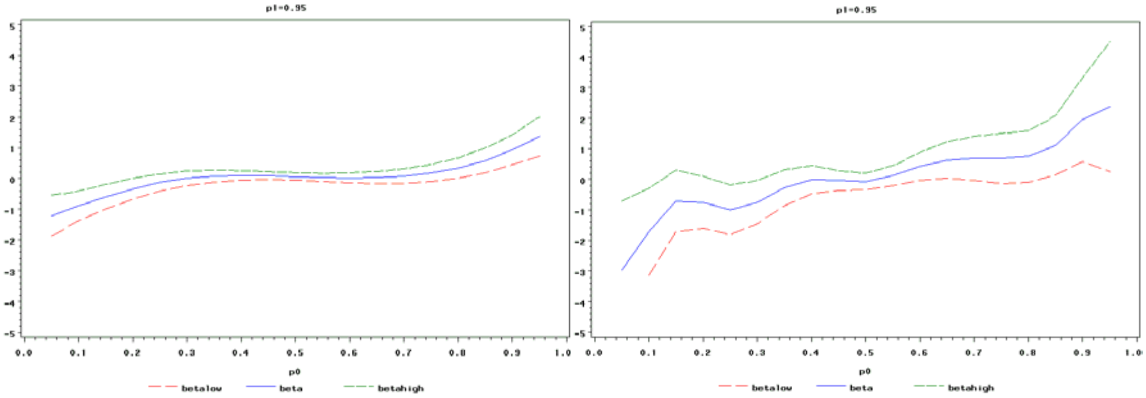

Figure 9. Marginal plot of the regression function on the percentile at time t, p_0 for the 95th percentile at time $t-1$. The number of knots used in the left graph was 3 and in the right 7. The three lines represent the estimated regression function and their 2-standard error intervals.