Journal of Forecasting
J. Forecast. **30**, 706–720 (2011)
Published online 8 December 2010 in Wiley Online Library
(wileyonlinelibrary.com) **DOI**: 10.1002/for.1204

Nonparametric Density Forecast Based on Time- and State-Domain

JOÃO NICOLAU*
School of Economics and Management (ISEG)/Technical University of Lisbon (UTL) and CEMAPRE, Lisbon, Portugal

ABSTRACT
We propose a new nonparametric density forecast based on time- and state-domain smoothing. We analyze some of its asymptotic properties and provide an empirical illustration.

KEY WORDS density forecasts; nonparametric methods; earnings forecast; time series

INTRODUCTION

Forecasting densities has been at the core of the finance and economic research agenda. For example, in risk management, it is fundamental to properly evaluate the density forecast of the change in the value of customized portfolios over a particular holding period. In general, most of the classical finance theories, such as portfolio selection, options valuation and asset pricing aim to model the uncertainty via distribution function. Also in economics there has been increasing interest in evaluating forecasting models of unemployment, inflation and output.

Consider a time series of cross-sections $X = \{X_{t,i}; t = 1, \ldots, T; i = 1, \ldots, n\}$ where for each t, $\{X_{t,i}; i = 1, \ldots, n\}$ is a random sample of n observations, and where $X_{t,i}$ may be correlated over time. Let $f_{t,i}(x)$ be the density of $X_{t,i}$ (at point x). We assume that $f_{t,i}(x) = f_{t,j}(x)$, $\forall i, j$, so we omit the subscript i in $f_{t,i}(x)$ and write $f_t(x)$ instead of $f_{t,i}(x)$. In our setting, $f_t(x)$ may be different from $f_s(x)$ either because X is nonstationary (for example, it may contain a unit root) or has a deterministic trend. Our framework is also suitable for the case where X is stationary but the conditional densities are time-dependent (for example, $X_{t,i}$ may follow an AR(p) or GARCH process).

To predict $f_{\tau+1}(\xi)$ we use (in principle) the past values of $f_t(\xi)$, $t \leq \tau$. This is plausible if there is some kind of autocorrelation in the sense $\text{cov}(h_1(X_{t,i}), h_2(X_{s,i})) \neq 0$, for any h_1 and h_2 real functions. For example, suppose $\text{cov}(|X_{t,i}|, |X_{t-1,i}|) > 0$, as is common to observe in financial markets (this measure is related to the so-called volatility clustering under which, as noted by Mandelbrot, 'large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes'). Under these circumstances, one should expect that the sequence of densities $\{f_t\}$, conditioned on a set of random variables, are (in some way) correlated. This point is illustrated with

* Correspondence to: João Nicolau, School of Economics and Management (ISEG)/Technical University of Lisbon (UTL) and CEMAPRE, ISEG, Rua do Quelhas 6, 1200-781 Lisbon, Portugal. E-mail: nicolau@iseg.utl.pt

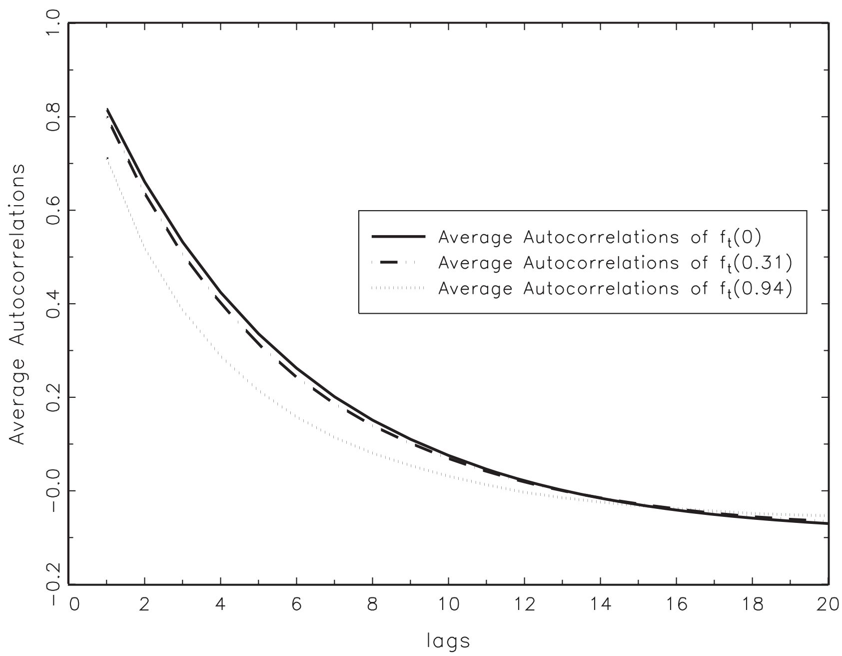

Figure 1. Autocorrelations of a density of GARCH process

three examples. In Figure 1 we plot the average autocorrelations of over 10,000 simulations of the conditional density of a GARCH process for various values of ξ. Specifically, we considered the GARCH process $y_t = \sigma_t\varepsilon_t$, where $\varepsilon_t \sim$ i.i.d.$N(0, 1)$ and $\sigma_t^2 = 0.01 + 0.3y_{t-1}^2 + 0.6\sigma_{t-1}^2$. The conditional density is $f_t(\xi) = (2\pi\sigma_t^2)^{-1/2}e^{-\xi^2/(2\sigma_t^2)}$. We analyzed three values for ξ: 0, $2\times\sqrt{\mathrm{var}(y_t)}$ and $3\times\sqrt{\mathrm{var}(y_t)}$ where $\mathrm{var}(y_t) = 0.01/(1 - 0.3 - 0.6)$ is the unconditional variance of y_t. Figure 1 clearly shows that the nonlinear correlation in y_t (for example, $|y_t|$ or y_t^2 exhibit strong autocorrelations) translates into autocorrelation in f_t (although y_t is not autocorrelated). The autocorrelations of f_t are higher when ξ is zero (i.e., when ξ coincides with the conditional mean, where the density reaches its maximum). Nevertheless, the autocorrelations associated with the other values of ξ are also relatively high. We also verified (results not shown) that the more persistent the GARCH process is, the higher the autocorrelations of f_t are.

Another example, from a nonstationary process involving a linear trend is presented in Figure 2. Part A1 shows a simulated path from the process $y_t = 0.005t + \varepsilon_t$, where $\varepsilon_t \sim$ i.i.d.$N(0, 1)$. Part A2 shows the corresponding sequence of marginal densities $\{f_t(\xi)\}$ across time. Although, in this case, it does not make sense to invoke the autocorrelation function of the densities (as they are deterministic), the point is that past values of f_t can be used to predict future values of f_t.

We provide a third example of an autocorrelated process with correlated densities. In this case we take an analytical approach. Let $y_t = \phi y_{t-1} + \varepsilon_t$, $|\phi| < 1$, where $\varepsilon_t \overset{\text{i.i.d.}}{\sim} N(0, 1)$ Let $f_t(\xi) \equiv f(\xi|y_{t-1})$ be the one-step-ahead conditional density (notice that we interpreted $f_t(\xi) \equiv f_t(\xi|y_{t-1})$ as a random variable, since it depends on y_{t-1} which, in turn, is interpreted as a random variable and not as the value of y observed at time $t - 1$). Now, it can be proved that

DOI: 10.1002/for

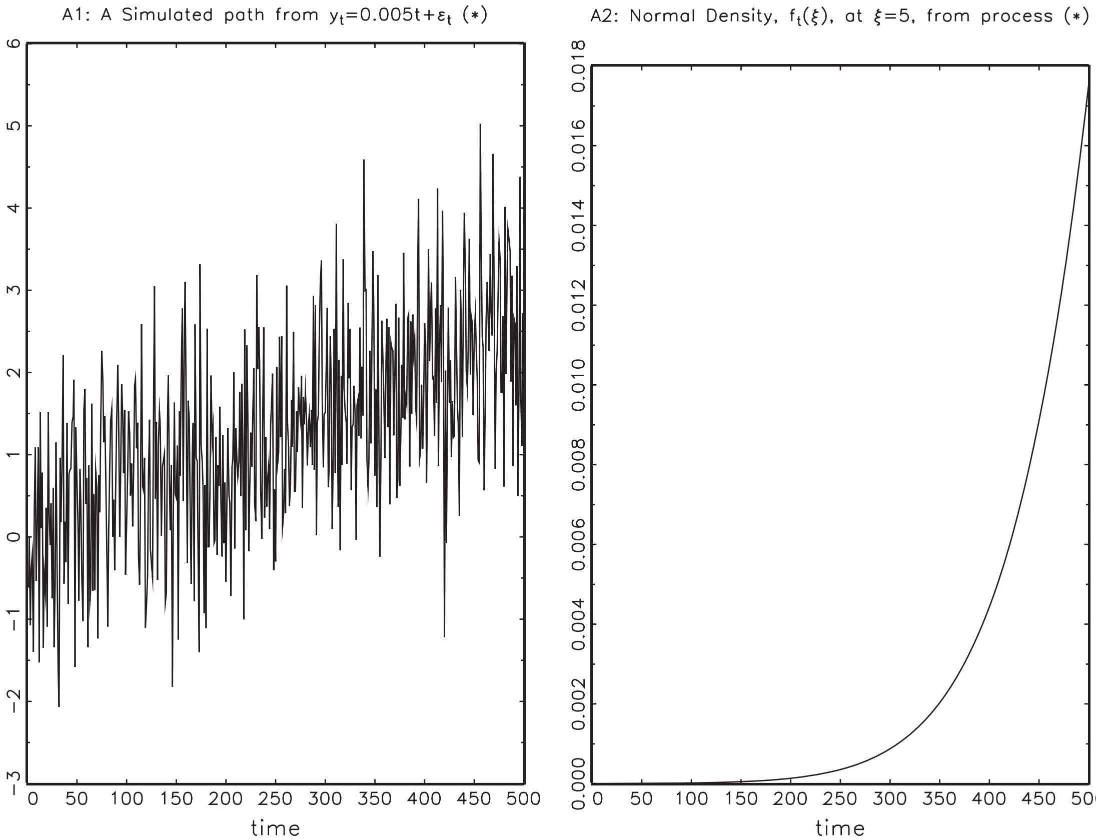

Figure 2. Simulated path from a nonstationary process and the corresponding densities (past values of f_t may be used to predict future values of f_t)

$$
\begin{aligned}
\operatorname{cov}(f_t(\xi), f_{t-1}(\xi)) &= \iint f_t(\xi|u)\, f_t(\xi|s)\, f_j(u,s)\, \mathrm{d}s\mathrm{d}u - \left(\int f_t(\xi|u)\, f_m(u)\, \mathrm{d}u\right)^2 \\
&= \frac{(\phi^2-1)}{2\pi}\left(\frac{e^{-\frac{\xi^2}{1+\phi^2}}}{1+\phi^2} - e^{-\frac{\xi^2\left(2+\phi^2(2+(\phi-2))\phi\right)}{2+4\left(\phi^2+\phi^4\right)}}\sqrt{\frac{1}{1+\phi^2-2\phi^6}}\right)
\end{aligned}
$$

where ($f_j(u, s)$ is the joint probability density function and f_m is the marginal probability density function; the proof is available upon request). Thus, when $\phi \neq 0$, i.e., when y is autocorrelated, the conditional densities are also autocorrelated. The message of the last paragraphs is clear: it makes sense to predict $f_{\tau+1}(\xi)$ based on the past values of $f_t(\xi)$, $t \leq \tau$.

The rest of the paper is organized as follows. In the next section we introduce the proposed estimator. In the third section we analyze some asymptotic properties. The fourth section deals with the estimation of α (we will see that the proposed estimator involves the 'nuisance' parameter α, that

DOI: 10.1002/for

needs to be estimated). The fifth section analyzes the bandwidth selection issue. The sixth section illustrates the method and finally, in the seventh section, we present some extensions that can be considered for future work.

A PREDICTOR FOR THE PROBABILITY DENSITY FUNCTION

To predict $f_{\tau+1}(\xi)$, using the past values of $f_\tau(\xi)$, $t \le \tau$, we propose the predictor

$$g_{\tau+1}(\xi) = \sum_{t=1}^{\tau} f_t(\xi)\omega_t, \quad \omega_t = \frac{\alpha^{\tau-t}(1-\alpha)}{1-\alpha^\tau}, \quad 0 \le \alpha \le 1$$

Figure 3 illustrates the idea. In this example the aim is to predict $f_6(\xi)$. Using the five past values of f_t (i.e., $f_1(\xi), \ldots, f_5(\xi)$), we form the sum $\sum_{t=1}^{5} f_t(\xi)\omega_t$. The most recent value $f_5(\xi)$ is weighted by $\omega_5 = (1 - \alpha)/(1 - \alpha^5)$ and the oldest value is weighted by $\omega_1 = \alpha^{5-1}(1 - \alpha)/(1 - \alpha^5) < \omega_5$.

Clearly, $g_{\tau+1}(\xi)$ is a weighted mean. The weights $\omega_t's$ decay exponentially fast as we move backwards in time. The result is that: (i) $\sum_{t=1}^{\tau} \omega_t = 1$; (ii) $\lim_{\alpha\to 0}\omega_t = 0$ if $t < \tau$, $\lim_{\alpha\to 0}\omega_t = 1$ if $t = \tau$; and (iii) $\lim_{\alpha\to 1}\omega_t = 1/\tau$. Thus, if $\alpha = 0$, $g_{\tau+1}(\xi)$ equals $f_\tau(\xi)$, if $\alpha = 1$, each value of f_t contributes equally to the $g_{\tau+1}(\xi)$ average.

In practice, we need to replace the unknown value $f_t(\xi)$ by the nonparametric estimate

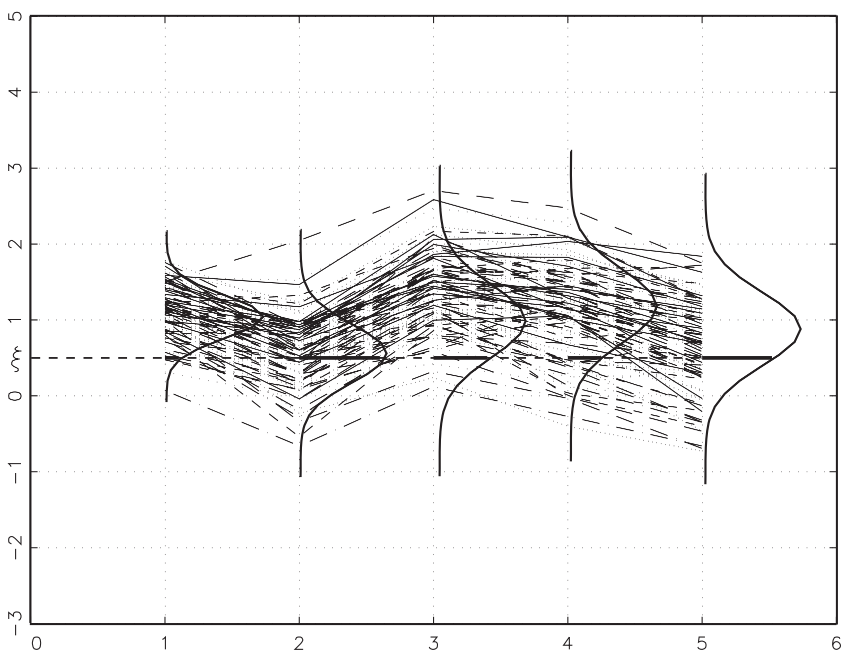

Figure 3. $f_t(\xi)$ is evaluated at t = 1, 2, 3, 4, 5 in a panel dataset

DOI: 10.1002/for

$$\hat{f}_t(\xi) = \sum_{i=1}^{n} \frac{1}{nh_t} K\left(\frac{X_{t,i} - \xi}{h_t}\right)$$

(the estimation of α is addressed in the fourth section). Hence the estimator of g is given by

$$\hat{g}_{\tau+1}(\xi) = \sum_{t=1}^{\tau} \hat{f}_t(\xi)\omega_t = \sum_{t=1}^{\tau}\sum_{i=1}^{n} \frac{1}{nh_t} K\left(\frac{X_{t,i} - \xi}{h_t}\right) \frac{\alpha^{\tau-t}(1-\alpha)}{1-\alpha^{\tau}} \tag{1}$$

Equation (1) allows us to see that $\hat{g}_{\tau+1}(\xi)$ uses smoothing techniques in both time and state domains. To exemplify, suppose that K is the Gaussian kernel. In this case, we have

$$K\left(\frac{X_{t,i} - \xi}{h_t}\right) \frac{\alpha^{\tau-t}(1-\alpha)}{1-\alpha^{\tau}} = \frac{1}{\sqrt{2\pi h_t}} e^{-\frac{(X_{t,i}-\xi)^2}{2h_t}} \alpha^{\tau-t} \frac{(1-\alpha)}{1-\alpha^{\tau}}$$

It is clear now that the $X_{t,i}$ value receives more weight (on the average $\hat{g}_{\tau+1}(\xi)$) when $X_{t,i}$ is closer to ξ and t is closer τ. Figure 4 illustrates the idea for $\xi = 0$, $\tau = 20$ and $\alpha = 0.8$ and $\alpha = 0.7$. To predict $f_{\tau+1}(0)$ at time $\tau + 1 = 21$, the most important values of the sample $X = \{X_{t,i};\ t = 1, \ldots, T;\ i = 1, \ldots, n\}$, on the average $\hat{g}_{\tau+1}(\xi)$, are those that are at the vicinity of $\xi = 0$ and $t = 20$.

The expression $g_{\tau+1}$ coincides with the true one-step-ahead density $f_{\tau+1}$ if one assumes that $f_{\tau+1}$ can be expressed through a weighted mean of past densities. If we are not willing to assume such a parametric specification, we may see $g_{\tau+1}$ as an approximation of $f_{\tau+1}$ whose quality may be assessed through the quantity $Q_{T,n}(\hat{\alpha})$, given in the fourth section. This is almost like a filtering problem where $f_{\tau+1}$ may be understood as the 'signal process' that cannot be observed directly. Given the observations X, we extract the 'signal' $f_{\tau+1}$ based on $\hat{g}_{\tau+1}(\xi;\ \alpha)$ and on the criterion $\min_{\alpha} Q_{T,n}(\alpha)$. As in the filtering problem, the solution $\hat{g}_{\tau+1}$ satisfies a criterion of optimally, but does not have to coincide with the signal process.

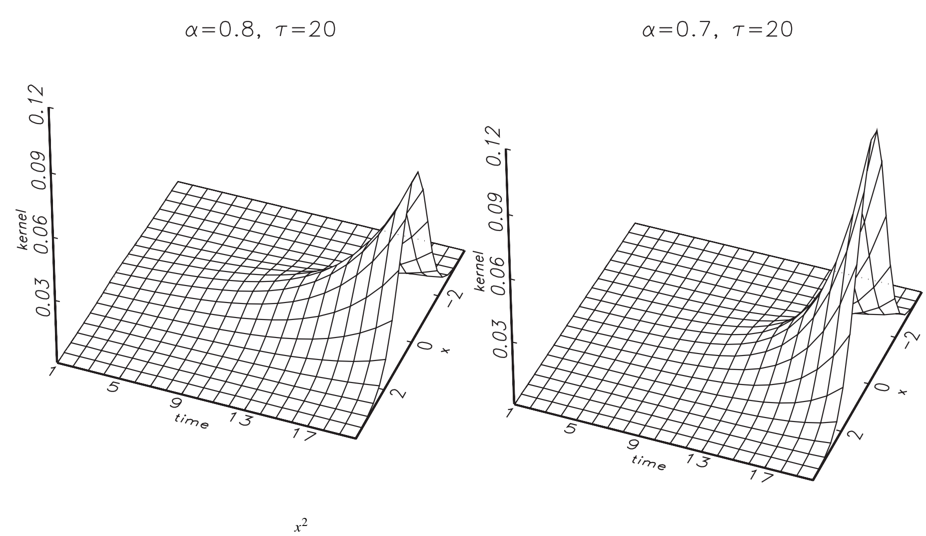

Figure 4. Kernel $K(x,t) = (2\pi)^{-1/2} e^{-\frac{x^2}{2}} \alpha^{\tau-t}(1-\alpha)/(1-\alpha^{\tau})$. Smoothing techniques are used in both time and state domain

SOME LIMIT RESULTS

Our limit results involve large cross-section-dimension data, $n \to \infty$, and fixed T. These hypotheses are plausible in our dataset, facilitate the proofs and avoid imposing unnecessary restrictions on the probabilistic behavior of X. We further discuss this issue in the sixth section.

We consider the following assumptions:

A1. $X_{t,i}$ and $X_{s,j}$ are independent for $i \neq j$.
A2. The kernel K is a symmetric function around zero satisfying: $\int K(u)\mathrm{d}u = 1$, $\int u^2 K(u)\mathrm{d}u < \infty$, $\int K^2(u)\mathrm{d}u < \infty$, $|u||K(u)| \to 0$ as $|u| \to \infty$, $\sup|K(u)| < \infty$, $\int K^{2+\delta}(u)\mathrm{d}u < \infty$ for $\delta > 0$.
A3. f is continuous, $\int|f(x)|\mathrm{d}x < \infty$, the second-order derivative of f is continuous and bounded in some neighborhood of ξ.
A4. $h_t \to 0$ as $n \to \infty$ and $nh_t \to \infty$ as $n \to \infty$.

These assumptions are very weak. In particular, stationarity is not imposed and no reference is made concerning the dependence of X (for example, a unit root is allowed). Under Assumptions A1–A4, we have the following results:

Theorem 1. $\hat{f}_t(\xi) \xrightarrow{p} f_t(\xi)$ and $\hat{g}_{\tau+1}(\xi) \xrightarrow{p} g_{\tau+1}(\xi)$.

Theorem 2. $\sqrt{nh}\left(\hat{g}_{\tau+1}(\xi) - \mathrm{E}(\hat{g}_{\tau+1}(\xi))\right) \xrightarrow{d} N(0, \sigma^2)$ where

$$\begin{aligned}\sigma^2 &= \lim_{n\to\infty} \operatorname{var}\left(\sqrt{nh}\sum_{t=1}^{\tau} f_t(\xi)\omega_t^2\right)\\ &= \int K^2(u)\,du\left(\sum_{t=1}^{\tau}\omega_t^2 f_t(\xi) + 2\sum_{t=1}^{\tau}\sum_{s=1}^{t-1}\omega_t\omega_s\rho_{(t-s)}f_t(\xi)\right), \qquad (2)\\ \rho_{(t-s)} &= \operatorname{cov}\left(K\left(\frac{X_{t,i}-\xi}{h_t}\right), K\left(\frac{X_{s,i}-\xi}{h_s}\right)\right)\Big/\operatorname{var}\left(K\left(\frac{X_{t,i}-\xi}{h_t}\right)\right)\end{aligned}$$

The asymptotic variance defined in equation (2) involves some unknown parameters. We may use the *plug-in-principle* and replace the unknown parameters with consistent estimates. However, the substitution of ω_t by a $\hat{\omega}_t$ estimate may alter the asymptotic distribution of $\hat{g}$ (cf. Theorem 2), as ω_t is a function of α and $\hat{g}$ depends explicitly on this α parameter. We analyze this issue in the next section.

ESTIMATION OF α

We now address the estimation of α. To obtain an estimate of α for each ξ we consider the following minimization problem: $\min_\alpha Q_{T,n}(\alpha)$, where

$$Q_{T,n}(\alpha) = \frac{1}{T}\sum_{\tau=0}^{T-1}\left(\hat{g}_{\tau+1}(\xi;\alpha) - \hat{f}_{\tau+1}(\xi)\right)^2$$

We write $\hat{g}_{\tau+1}(\xi;\ \alpha)$ instead of $\hat{g}_{\tau+1}(\xi)$ to emphasize the dependency of $\hat{g}$ on α. The idea with this optimization problem is simple: we look for the value of α that minimizes the mean square error of the forecast. Notice that the smoothing parameter α is (generally) different depending

DOI: 10.1002/for

on the value of ξ. We assume that there is a unique value $\alpha_0 \in [0, 1]$ such that $\frac{1}{T}\sum_{\tau=0}^{T-1}\left(\sum_{t=1}^{\tau} f_t(\xi)\frac{\alpha^{\tau-t}(1-\alpha)}{1-\alpha^{\tau}} - f_{\tau+1}(\xi)\right)^2$ is a minimum. This assumption is violated if all $f_t(\xi)$, $t = 1, \ldots, T-1$ are equal. Thus we implicitly assume that at least one density f_t is different from the other densities. Let $\hat{\alpha} = \operatorname{argmin}_{\alpha\in[0,1]} Q_{T,n}(\alpha)$. We have the following results.

Theorem 3. $\hat{\alpha} \xrightarrow{p} \alpha_0$.

Theorem 4. $\sqrt{nh}\left(\hat{g}_{\tau+1}(\xi;\hat{\alpha}) - \mathrm{E}(\hat{g}_{\tau+1}(\xi;\alpha))\right) \xrightarrow{d} N(0,\sigma^2)$, where σ^2 is given in Theorem 2.

Hence $\sqrt{nh}\left(\hat{g}_{\tau+1}(\xi;\hat{\alpha}) - \mathrm{E}(\hat{g}_{\tau+1}(\xi;\alpha))\right)$ behaves the same way asymptotically whether we use $\hat{\alpha}$ or α.

BANDWIDTH SELECTION

We will briefly discuss the bandwidth selection problem. There are several strategies in dealing with bandwidths $\{h_t\}$. One strategy consists of assuming that the bandwidths are equal, say $h = h_1 = \ldots = h_n$. In this case, one may use as a criterion to be optimized the asymptotic mean integrated squared error (AMISE). Given the proofs in the Appendix, one obtains

$$\text{AMISE} = \frac{\int K^2(u)\,\mathrm{d}u\left(1 + 2\sum_{t=1}^{\tau}\sum_{s=1}^{t-1}\omega_t\omega_s\rho_{(t-s)}\right)}{nh} + \frac{h^4}{4}\int\left[\sum_{t=1}^{\tau}\left(\int u^2 K(u)\,\mathrm{d}u\right) f_t''(x)\,\omega_t\right]^2 \mathrm{d}x$$

The solution of $\min_h \text{AMISE}$ is $h^* = cn^{-1/5}$, where c is a complicated expression involving ω_t, $\rho_{\rho(t-s)}$ and $f_t''(x)$. Different approaches can be considered in order to deal with c. For example, one may choose a reference density for f_t (e.g., normal density); or one may consider the *plug-in-principle* to deal with expressions such as $\rho_{(t-s)}$, $\int f_t''(x)\mathrm{d}x$ and $\int f_t''(x) f_s''(x)\mathrm{d}x$, etc. A deeper analysis of this topic is beyond the scope of the present paper.

On the other hand, if one assumes that, in fact, the bandwidths vary across time we may apply the existing methods of bandwidth selection to obtain the 'optimal' h_t for each density estimation $\{\hat{f}_t(\xi), t = 1, \ldots, \tau + 1\}$. An alternative consists of finding the 'optimal' h_t using a criterion based on forecast errors such as

$$\min_{\alpha,h_t>0} Q_{T,n}(\alpha,h_t) = \min_{\alpha,h_t>0}\frac{1}{T}\sum_{\tau=0}^{T-1}\left(\sum_{t=1}^{\tau}\sum_{i=1}^{n}\frac{1}{nh_t}K\left(\frac{X_{t,i}-\xi}{h_t}\right)\frac{\alpha^{\tau-t}(1-\alpha)}{1-\alpha^{\tau}} - \hat{f}_{\tau+1}(\xi)\right)^2 \qquad (3)$$

The bandwidth associated with $\hat{f}_{\tau+1}(\xi)$ may be based on some of the existing methods, whereas $\{h_t, t = 1, \ldots, \tau\}$ solve the above optimization problem. Hence, these optimal bandwidths $\{h_t, t = 1, \ldots, \tau\}$ are chosen in order to obtain the best forecasts in L^2 sense. In the next section, we use this principle to obtain the bandwidths, but considering $h_t = \eta_t(4/3)^2\hat{\sigma}X_t n^{-1/5}$ so that the minimization is done with respect to α and η_t. Obviously, the optimization problems $\min_{\alpha,h_t>0} Q_T, n(\alpha, h_t)$ and $\min_{\alpha,\eta_t>0} Q_T, n(\alpha, \eta_t)$ are equivalent. Notice that the estimates η_t may be compared with the

DOI: 10.1002/for

bandwidth $h_t^{\text{normal}} = (4/3)^2 \hat{\sigma} X_t n^{-1/5}$, which is optimal when $X_{t,i}$ has normal distribution and the Kernel is Gaussian.

EMPIRICAL ILLUSTRATION

To illustrate the method we consider the same data as Pesaran (2007), which is available at the site http://qed.econ.queensu.ca/jae/datasets/pesaran002/. The data are the real earnings of households and were drawn from the 1971–1992 family and individual-merged files of the Panel Study of Income Dynamics (longitudinal survey of a nationally representative sample of US families begun in 1968). As reported in Pesaran (2007), the data were defined to include all households with male heads aged 25–55 with 22 years of usable earnings. As a result, the dataset is of 22×181 observations (22 years; 181 households). Given the short interval of time of our dataset, we consider nine forecasts out-of-sample. We start using observations from 1971 to 1983 and then obtain the forecast for the density in 1984. The estimated predictor is $\hat{g}_{\tau+1}$, where $\tau + 1$ refers to the year 1984. This forecast is then compared with the target density of the year 1984, $\hat{f}_{\tau+1}$, obtained using all the observations of that year. Thus the quantity $|\hat{g}_{\tau+1} - \hat{f}_{\tau+1}|$ gives a measure of the forecast precision. The quality of the proposed forecasts is also assessed by comparing $\hat{g}_{\tau+1}$ to a kernel estimation based on a method suggested by an anonymous referee. It consists of forecasting $X_{t+1,i}$ for fixed i, using exponential smoothing over $X_{t,i}$ and lags, and then building the density forecasts over the forecasts $\hat{X}_{t+1,i}$ using kernel estimation. We also considered different smoothing parameters for each i, which were obtained by minimizing the one-step-ahead forecast errors between $\hat{X}_{\tau+1,i}$ and $X_{\tau+1,i}$. Let us denote this kernel density by $\hat{h}_{\tau+1}(\xi) = 1/(nh)\sum_{i=1}^{n} K\left(\left(\hat{X}_{\tau+1,i} - \xi\right)/h\right)$. To get a picture of the entire density we select various ξ_j values across the most relevant interval of state space of X. We use 50 equidistant ξ_j values according to the rule $\min_{t,i} X_{t,i} < \xi_1 < \xi_2 < \ldots < \xi_{50} < \max_{t,i} X_{t,i}$.

Next, we use the observations from 1971 to 1984 to obtain the forecast for the density in 1985. The estimates $\hat{g}$, $\hat{f}$ and $\hat{h}$ are recalculated using the new information. This procedure is repeated until the forecast of the last year is obtained. In this application we consider the Gaussian kernel and for simplicity we use the bandwidth $h_t = (4/3)^2 \sigma_{X_t} n^{-1/5}$, which is optimal when $X_{t,i}$ has normal distribution. The results are presented in Figure 5. Both estimators seem to be close to the density $\hat{f}$, but the best results are obtained using the proposed estimator $\hat{g}$. To get a more precise measure of the estimators' quality, we computed the empirical integrated mean square errors, that is, the mean of $(\hat{g}_{\tau+1}(\xi_j) - \hat{f}_{\tau+1}(\xi_j))^2$ and $(\hat{h}_{\tau+1}(\xi_j) - \hat{f}_{\tau+1}(\xi_j))^2$ over the set $\{\xi_1, \ldots, \xi_{50}\}$ and over the nine forecasts out-of-sample. The empirical integrated mean square errors of $\hat{g}$ and $\hat{h}$ are 0.00068 and 0.00179 respectively. The general conclusion is that the proposed estimator performs very well and better than the estimator $\hat{h}_{\tau+1}$. Although $\hat{h}_{\tau+1}$ is very intuitive and easily computed, it assumes a parametric linear representation of the underlying process, which may be a disadvantage if the true process is nonlinear. I think the estimator $\hat{h}$ can be improved in several ways (e.g., the smoothing parameter may be chosen to minimize the integrated mean square errors; another possibility is to consider a different model to forecast $\hat{X}_{\tau+1,i}$).

When the bandwidths are selected according to the optimization problem (3) the results are even better. Using the whole sample, we found the ratio $\sqrt{\hat{Q}_{T,n}(\alpha)} \Big/ \sqrt{\hat{Q}_{T,n}(\alpha, \eta_t)} = 1.33$. This means that the root mean square error of the forecast using the simple rule $h_t = (4/3)^2 \hat{\sigma}_{X_t} n^{-1/5}$ is about 33% higher than that of the forecast using the 'optimal' bandwidths, obtained in the context of minimization problem (3). These results are expected in view of the fact that these 'optimal' bandwidths are chosen in order to obtain the best forecasts in L^2 sense.

DOI: 10.1002/for

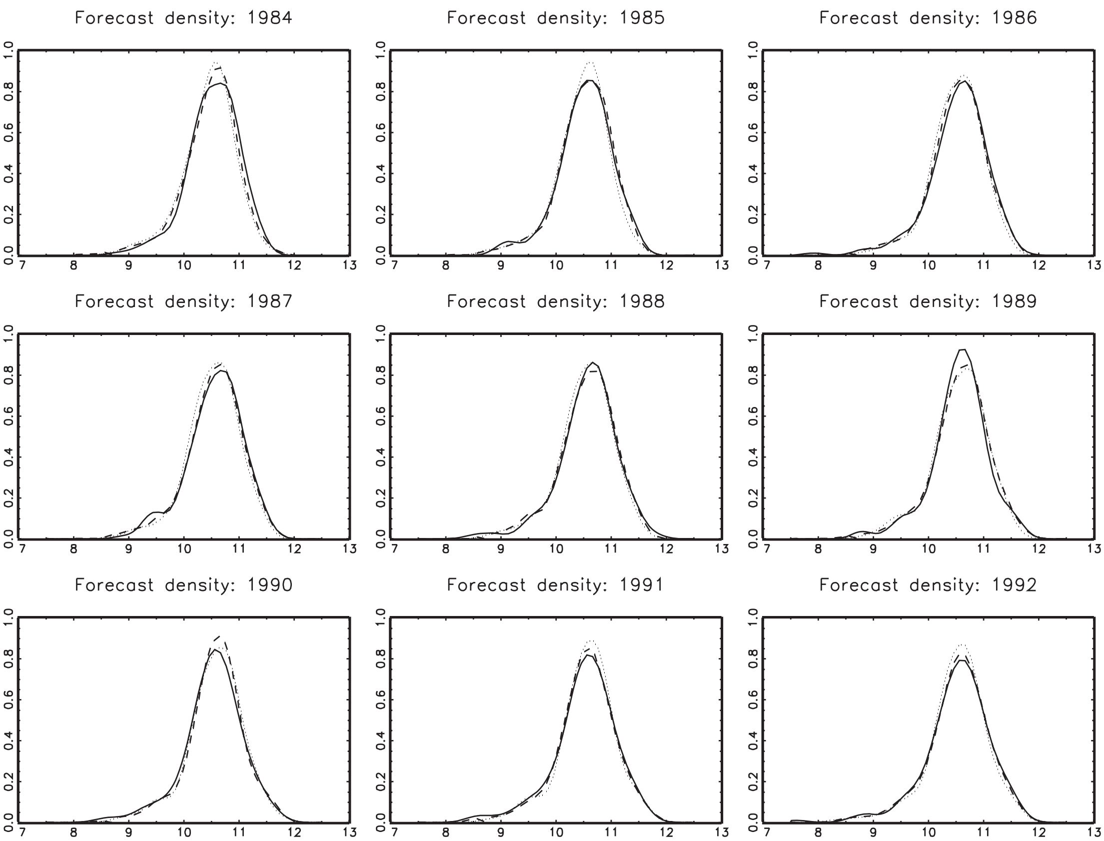

Figure 5. Assessing the quality of density forecast. Legend: — f; - - - - g; . . . h

EXTENSIONS

There are a number of extensions that could be interesting to investigate in the future.

- The ideas presented in this work can be immediately applied to conditional moment estimation and nonparametric estimation of derivatives, to name only two areas.
- The case $n,T \to \infty$ can also be studied. When analyzing $T \to \infty$, temporal dependence is generally a crucial issue that needs to be accounted for. However, temporal dependence is not, in principle, a major concern in the context of our estimator as the $\omega_t' s$ weights naturally impose an exponential decay in the autocorrelation of $\hat{f}_t$, whether or not X is strongly dependent (see our proofs in the Appendix). For this reason, our results are robust for large T.
- One may also consider the case where the vector $\{X_{t,1}, \ldots, X_{t,n}\}$ for fixed t, is interpreted as describing n characteristics of a given population. In this case, cross-section correlations must be taken into consideration.
- A way to obtain $\hat{\alpha}$ would be by minimizing $\sigma^2 = \sigma^2(\alpha)$ with respect to α (see equation (2)). My conjecture is that this method is inferior to the one proposed in this paper.

- The minimization problem $\min_\alpha Q_{T,n}(\alpha)$ can be seen as a minimization of the mean square error between the $\hat{f}_{\tau+1}(\xi)$ and $\hat{g}_{\tau+1}(\xi;\alpha)$. It turns out that this mean square error may be evaluated across different values of ξ. Actually, the same principle is behind the mean integrated square error. Therefore, the α parameter may be also obtained through the minimization problem:

$$\min_\alpha \sum_{\tau=0}^{T-1}\sum_{i=1}^{N}\left(\hat{f}_{\tau+1}(\xi_i)-\hat{g}_{\tau+1}(\xi_i;\alpha)\right)^2$$

This method has one disadvantage, however. It implies using the same estimate of α across all the ξ_i values. In the proposed method, we have an α estimate for each ξ_i point evaluated. It is quite natural to expect a more precise estimate when α is allowed to vary according to the value of ξ.

APPENDIX: PROOFS

Lemma 1. Let $\omega_t = \alpha^{\tau-t}(1-\alpha)/(1-\alpha^\tau)$ with $0 \le \alpha \le 1$ and z_t a real function of t. Then

$$\left|\sum_{i=1}^{\tau}\omega_t z_t\right| \le C_1 z, \quad \left|\sum_{i=1}^{\tau}\omega_t^2 z_t\right| \le C_2 z \qquad \left|\sum_{i=1}^{\tau}\sum_{i=1}^{t-1}\omega_t\omega_s z_t\right| \le C_3 z$$

where $0 < C_i < \infty$, $i = 1, 2, 3$ and $z = \max_t\{z_t\}$. These results still hold when $\tau \to \infty$.

The proof is straightforward. Notice that $\lim_{\alpha\to 0}\omega_t = 0$ if $t < \tau$ and $\lim_{\alpha\to 0}\omega_t = 1$ if $t = \tau$ and $\lim_{\alpha\to 1}\omega_t = 1/\tau$.

Remark on the proofs. In the following proofs we assume, without any loss of generality, that the bandwidths do not depend on t, i.e., $h_1 = \ldots = h_t = h$. This greatly simplifies the proofs. For example, to discuss $\sum_{t=1}^{\tau} O(h_t)\omega_t$, without assuming this simplification, we would have to consider

$$\left|\sum_{t=1}^{\tau}O(h_t)\omega_t\right| \le \left|\sum_{t=1}^{\tau}O(\max\{h_t\})\omega_t\right| = |O(\max\{h_t\})|\left|\sum_{t=1}^{\tau}\omega_t\right| = |O(\max\{h_t\})|$$

Thereby, $\left|\sum_{t=1}^{\tau} O(h_t)\omega_t\right| \to 0$ as $\max_t\{h_t\} \to 0$. We avoid all this reasoning by simply assuming that $h_1 = \ldots = h_t = h$, as the final result is the same whether or not h is time-dependent, in view of Assumption A4: $h_t \to 0$ as $n \to \infty$ and $nh_t \to \infty$ as $n \to \infty$. With $h_1 = \ldots = h_t = h$ the calculation is easier: $\sum_{t=1}^{\tau} O(h_t)\omega_t = O(h)\sum_{t=1}^{\tau}\omega_t = O(h)$.

Proof of Theorem 1. The result $\hat{f}_t(\xi) \xrightarrow{p} f_t(\xi)$ is well known, so we omit the proof. Let $\omega_t = \alpha^{\tau-t}(1-\alpha)/(1-\alpha^\tau)$. Under the assumptions of the theorem we have $\mathrm{E}(\hat{f}_t(\xi)) = f_t(\xi) + O(h^2)$ (see, for example, Pagan and Ullah, 1999, ch. 2). Hence

$$\begin{aligned}\mathrm{E}(\hat{g}_{\tau+1}(\xi)) &= \mathrm{E}\left(\sum_{t=1}^{\tau}\hat{f}_t(\xi)\omega_t\right) = \sum_{t=1}^{\tau}\mathrm{E}\left(\hat{f}_t(\xi)\right)\omega_t = \sum_{t=1}^{\tau}\left(f_t(\xi)+O(h^2)\right)\omega_t \\ &= g_{\tau+1}(\xi) + O(h^2)\sum_{t=1}^{\tau}\omega_t \to g_{\tau+1}(\xi), \quad \left(\sum_{t=1}^{\tau}\omega_t = 1\right)\end{aligned}$$

DOI: 10.1002/for

as $h \to 0$ (see previous remark). Now, consider

$$\begin{aligned} \operatorname{var}(\hat{g}_{\tau+1}(\xi)) &= \operatorname{var}\left(\sum_{t=1}^{\tau}\hat{f}_t(\xi)\omega_t\right) \\ &= \sum_{t=1}^{\tau}\operatorname{var}\left(\hat{f}_t(\xi)\omega_t\right) + 2\sum_{t=1}^{\tau}\sum_{s=1}^{t-1}\operatorname{Cov}\left(\hat{f}_t(\xi)\omega_t, \hat{f}_s(\xi)\omega_s\right) \end{aligned} \tag{4}$$

Firstly, let us focus on the first term of (4). We have

$$\begin{aligned} \sum_{t=1}^{\tau}\operatorname{var}\left(\hat{f}_t(\xi)\omega_t\right) &= \sum_{t=1}^{\tau}\omega_t^2\operatorname{var}\left(\hat{f}_t(\xi)\right) = \sum_{t=1}^{\tau}\omega_t^2\operatorname{var}\left(\sum_{i=1}^{n}\frac{1}{nh}K\left(\frac{X_{t,i}-\xi}{h}\right)\right) \\ &= \sum_{t=1}^{\tau}\omega_t^2\left((nh)^{-1}\int K^2(u)f_t(hu+\xi)\mathrm{d}u - n^{-1}\left(\int K(u)f_t(hu+\xi)\mathrm{d}u\right)^2\right) \\ &= \sum_{t=1}^{\tau}\omega_t^2\left(O\left(\frac{1}{nh}\right)+O\left(\frac{1}{n}\right)\right) \to 0 \end{aligned}$$

as $nh \to 0$ and $n \to \infty$ (see Lemma 1).

Let us analyze the second term of (4), $2\sum_{t=1}^{\tau}\sum_{s=1}^{t-1}\operatorname{cov}\left(\hat{f}_t(\xi)\omega_t, \hat{f}_s(\xi)\omega_s\right)$. Let

$$\rho_{(t-s)} = \operatorname{cov}\left(K\left(\frac{X_{t,i}-\xi}{h}\right), K\left(\frac{X_{s,i}-\xi}{h}\right)\right)\Big/\operatorname{var}\left(K\left(\frac{X_{t,i}-\xi}{h}\right)\right)$$

Firstly, observe that

$$\begin{aligned} \operatorname{cov}\left(\hat{f}_t(\xi), \hat{f}_s(\xi)\right) &= \operatorname{cov}\left(\sum_{i=1}^{n}\frac{1}{nh}K\left(\frac{X_{t,i}-\xi}{h}\right), \sum_{i=1}^{n}\frac{1}{nh}K\left(\frac{X_{s,i}-\xi}{h}\right)\right) \\ &= \left(\frac{1}{nh}\right)^2\sum_{i=1}^{n}\operatorname{cov}\left(K\left(\frac{X_{t,i}-\xi}{h}\right), K\left(\frac{X_{s,i}-\xi}{h}\right)\right) \\ &= \left(\frac{1}{nh}\right)^2 n\operatorname{cov}\left(K\left(\frac{X_{t,i}-\xi}{h}\right), K\left(\frac{X_{s,i}-\xi}{h}\right)\right) \\ &= \left(\frac{1}{nh}\right)^2 n\rho_{(t-s)}\operatorname{var}\left(K\left(\frac{X_{t,i}-\xi}{h}\right)\right) \\ &= \left(\frac{1}{nh}\right)^2 n\rho_{(t-s)}\left(h\int K^2(u)f_t(hu+\xi)\mathrm{d}u - h^2\left(\int K(u)f_t(hu+\xi)\mathrm{d}u\right)^2\right) \\ &= \frac{1}{nh}\rho_{(t-s)}\left(\int K^2(u)f_t(hu+\xi)\mathrm{d}u - O(h)\right) \end{aligned}$$

DOI: 10.1002/for

Hence

$$\sum_{t=1}^{\tau}\sum_{s=1}^{t-1}\operatorname{cov}\left(\hat{f}_t(\xi)\omega_t,\hat{f}_s(\xi)\omega_s\right)=\sum_{t=1}^{\tau}\sum_{s=1}^{t-1}\omega_t\omega_s\operatorname{cov}\left(\hat{f}_t(\xi),\hat{f}_s(\xi)\right)$$
$$=\sum_{t=1}^{\tau}\sum_{s=1}^{t-1}\omega_t\omega_s\frac{1}{nh}\rho_{(t-s)}\left(\int K^2(u)f_t(hu+\xi)\mathrm{d}u-O(h)\right)$$
$$=\frac{1}{nh}\sum_{t=1}^{\tau}\sum_{s=1}^{t-1}\omega_t\omega_s\rho_{(t-s)}\left(\int K^2(u)f_t(hu+\xi)\mathrm{d}u-O(h)\right)$$

Assuming the most unfavorable situation, i.e., $|\sigma_{(t-s)}|=1$, we have

$$\left|\sum_{t=1}^{\tau}\sum_{s=1}^{t-1}\operatorname{cov}\left(\hat{f}_t(\xi)\omega_t,\hat{f}_s(\xi)\omega_s\right)\right|\le\frac{1}{nh}\sum_{t=1}^{\tau}\sum_{s=1}^{t-1}|\omega_t\omega_s||O(1)|$$
$$=O\left(\frac{1}{nh}\right)\sum_{t=1}^{\tau}\sum_{s=1}^{t-1}\omega_t\omega_s=O\left(\frac{1}{nh}\right)\frac{\alpha(1+\alpha+\alpha^2)}{(1+\alpha)^2(1+\alpha^2)}\to 0$$

as $nh\to\infty$. In conclusion, $\mathrm{E}(\hat{g}_{\tau+1}(\xi))\to g_{\tau+1}(\xi)$ and $\operatorname{var}\left(\sum_{t=1}^{\tau}\hat{f}_t(\xi)\frac{\alpha^{\tau-t}(1-\alpha)}{1-\alpha^{\tau}}\right)\to 0$; therefore $\hat{g}_{\tau+1}(\xi)\xrightarrow{p}g_{\tau+1}(\xi)$. □

Proof of Theorem 2. It is well known that

$$\sqrt{nh}\left(\hat{f}_t(\xi)-\mathrm{E}\left(\hat{f}_t(\xi)\right)\right)\xrightarrow{d}N\left(0,f_t(\xi)\int K^2(u)\mathrm{d}u\right)$$

(see, for example, Pagan and Ullah, 1999). Since $\hat{g}_{\tau+1}(\xi)$ is a linear combination of $\hat{f}_{t's}$, with nonstochastic weights, it follows that

$$\sqrt{nh}\,(\hat{g}_{\tau+1}(\xi)-\mathrm{E}(\hat{g}_{\tau+1}(\xi)))\xrightarrow{d}N\left(0,\lim_{n\to\infty}\operatorname{var}\left(\sqrt{nh}\sum_{t=1}^{\tau}\hat{f}_t(\xi)\omega_t\right)\right)$$

From the proof of Theorem 2 we know that

$$\operatorname{var}\left(\sqrt{nh}\sum_{t=1}^{\tau}\left(\hat{f}_t(\xi)\omega_t\right)\right)=\sum_{t=1}^{\tau}\omega_t^2\left(\int K^2(u)f_t(hu+\xi)\mathrm{d}u-n^{-1}\left(\int K(u)f_t(hu+\xi)\mathrm{d}u\right)^2\right)+$$
$$2\sum_{t=1}^{\tau}\sum_{s=1}^{t-1}\omega_t\omega_s\rho_{(t-s)}\left(\int K^2(u)f_t(hu+\xi)\mathrm{d}u-O(h)\right)$$
$$=\sum_{t=1}^{\tau}\omega_t^2\int K^2(u)f_t(hu+\xi)\mathrm{d}u-\sum_{t=1}^{\tau}\omega_t^2n^{-1}\left(\int K(u)f_t(hu+\xi)\mathrm{d}u\right)^2+$$
$$2\sum_{t=1}^{\tau}\sum_{s=1}^{t-1}\omega_t\omega_s\rho_{(t-s)}\int K^2(u)f_t(hu+\xi)\mathrm{d}u-\frac{2}{nh}\sum_{t=1}^{\tau}\sum_{s=1}^{t-1}\omega_t\omega_s\rho_{(t-s)}O(h)$$

Now, Lemma 1 allows us to conclude, as $nh\to\infty$ and $h\to 0$,

$$\left|\sum_{t=1}^{\tau}\omega_t^2 n^{-1}\left(\int K(u)f_t(hu+\xi)\mathrm{d}u\right)^2\right| \leq n^{-1}\max_t\left\{\left(\int K(u)f_t(hu+\xi)\mathrm{d}u\right)^2\right\} \to 0,$$

$$\left|\frac{2}{nh}\sum_{t=1}^{\tau}\sum_{s=1}^{t-1}\omega_t\omega_s\rho_{(t-s)}O(h)\right| \leq \left|\frac{2}{nh}C_3 O(h)\right| \to 0$$

Hence, at the limit,

$$\lim \mathrm{var}\left(\sqrt{nh}\sum_{t=1}^{\tau}\left(\hat{f}_t(\xi)\omega_t\right)\right) = \sum_{t=1}^{\tau}\omega_t^2 f_t(\xi)\int K^2(u)\mathrm{d}u + 2\sum_{t=1}^{\tau}\sum_{s=1}^{t-1}\omega_t\omega_s\rho_{(t-s)}f_t(\xi)\int K^2(u)\mathrm{d}u$$
$$= \int K^2(u)\mathrm{d}u\left(\sum_{t=1}^{\tau}\omega_t^2 f_t(\xi) + 2\sum_{t=1}^{\tau}\sum_{s=1}^{t-1}\omega_t\omega_s\rho_{(t-s)}f_t(\xi)\right) \qquad \square$$

Proof of Theorem 3. Let the sample space Ω and the parameter space Θ be endowed with the σ-algebra $\mathcal{A}$ and $\mathcal{B}$ respectively. In our case, $\alpha \in \Theta = (0, 1]$. Consider the following classical assumptions. (1) The parameter space, say Θ, is compact; (2) the mapping $Q_{T,n}(\alpha) = Q_{T,n}(\omega, \alpha)$, where $\omega \in \Omega$, from $\Omega \times \Theta$ satisfies (i) $Q_{T,n}(\cdot,\alpha)$: $\omega \to Q(\omega, \alpha)$ is mensurable from $(\omega, \mathcal{A})$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ for every $\alpha \in \Theta$, (ii) the mapping $Q_{T,n}(\omega, \cdot)$: $\alpha \to Q(\omega, \alpha)$ is continuous for every $\omega \in \Omega$; (3) the $Q_{T,n}$ sequence converges in probability uniformly in α to $Q_{T,\infty}$; (4) the $Q_{T,\infty}$ function is such that the minimum $\min_{\alpha\in[0,1]} Q_{T,\infty}(\omega, \alpha)$ is attained at a unique value α_0, independent of ω. Under (1)–(4) we have $\alpha \xrightarrow{p} \alpha_0$ (see, for example, Gouriéroux and Monfort, 1995, ch. 24). In our problem, conditions (1)–(2) are immediately verified. We prove now that (3) still holds. Following Davidson (1994, theorem 21.9), $Q_{T,n}(\alpha)$ converges in probability to $Q_{T,n}(\alpha)$ uniformly on a set [0, 1] if and only if a) $Q_{T,n}(\alpha) \xrightarrow{p} Q_{T,\infty}(\alpha)$ for each $\alpha \in [0, 1]$ and b) $Q_{T,n}(\alpha)$ is stochastically equicontinuous. By Slutsky's theorem and the fact that T is constant we have pointwise convergence (plim stands for limit in probability):

$$\mathrm{plim}_{n\to\infty}Q_{T,n}(\alpha) = \mathrm{plim}_{n\to\infty}\frac{1}{T}\sum_{\tau=0}^{T-1}\left(\hat{f}_{\tau+1}(\xi) - \hat{g}_{\tau+1}(\xi;\alpha)\right)^2$$
$$= \frac{1}{T}\sum_{\tau=0}^{T-1}\left(\mathrm{plim}_{n\to\infty}\hat{f}_{\tau+1}(\xi) - \mathrm{plim}_{n\to\infty}\hat{g}_{\tau+1}(\xi;\alpha)\right)^2$$
$$= Q_{T,\infty}(\alpha) = \frac{1}{T}\sum_{\tau=0}^{T-1}\left(f_{\tau+1}(\xi) - \sum_{t=1}^{\tau}f_t(\xi)\frac{\alpha^{\tau-t}(1-\alpha)}{1-\alpha^{\tau}}\right)^2$$

(see Theorem 1) (notice that T is kept constant and $n \to \infty$). This proves (a) $Q_{T,n}(\alpha) \xrightarrow{p} Q_{T,\infty}(\alpha)$ for each $\alpha \in [0, 1]$. To prove (b) it is sufficient that $\sup_{\alpha\in[0,1]}|\mathrm{d}Q_{T,n}(\alpha)/\mathrm{d}\alpha|$ be bounded in probability (Davidson, 1994, theorem 21.10). This is condition is now verified:

$$\left|\frac{dQ_T(\alpha)}{d\alpha}\right| = \frac{2}{T}\left|\sum_{\tau=0}^{T-1}\left(\hat{f}_{\tau+1}(\xi) - \hat{g}_{\tau+1}(\xi;\alpha)\right)\frac{\partial\hat{g}_{\tau+1}(\xi;\alpha)}{\partial\alpha}\right|$$
$$\leq \frac{2}{T}\left(\sum_{\tau=0}^{T-1}\left|\hat{f}_{\tau+1}(\xi) - \hat{g}_{\tau+1}(\xi;\alpha)\right|^2\right)^{1/2}\left(\sum_{\tau=0}^{T-1}\left|\frac{\partial\hat{g}_{\tau+1}(\xi;\alpha)}{\partial\alpha}\right|^2\right)^{1/2}$$

The first term $\frac{2}{T}\left(\sum_{\tau=0}^{T-1}\left|\hat{f}_{\tau+1}(\xi)-\hat{g}_{\tau+1}(\xi;\alpha)\right|^2\right)^{1/2}$ is finite because with T constant and $n \to \infty$. In fact, $\left|\hat{f}_{\tau+1}(\xi)-\hat{g}_{\tau+1}(\xi;\alpha)\right| \xrightarrow{p} |f_{\tau+1}(\xi)-g_{\tau+1}(\xi;\alpha)| < \infty$ and $\left|\hat{f}_{\tau+1}(\xi)-\hat{g}_{\tau+1}(\xi;\alpha)\right|^2 \xrightarrow{p} \left|\hat{f}_{\tau+1}(\xi)-\hat{g}_{\tau+1}(\xi;\alpha)\right|^2 \xrightarrow{p} |f_{\tau+1}(\xi)-g_{\tau+1}(\xi;\alpha)|^2 < \infty$ by (Slutsky's theorem). On the other hand

$$\sum_{\tau=0}^{T-1}\left|\frac{\partial \hat{g}_{\tau+1}(\xi;\alpha)}{\partial\alpha}\right|^2 = \sum_{\tau=0}^{T-1}\hat{f}_t^2\left(\frac{\mathrm{d}\omega(t)}{d\alpha}\right)^2 \xrightarrow{p} \sum_{\tau=0}^{T-1} f_t^2\left(\frac{\mathrm{d}\omega(t)}{d\alpha}\right) < \infty$$

(see Theorem 1), for all $\alpha \in [0, 1]$. Hence $|\mathrm{d}Q_T(\alpha)/\mathrm{d}\alpha| = O_p(1)$. In conclusion, $Q_{T,n}(\alpha)$ converges in probability to $Q_{T,n}(\alpha)$ uniformly on a set [0, 1]. Finally, by assumption $\alpha_0 = \operatorname{argmin} Q_{T,\infty}(\alpha)$. □

Proof of Theorem 4. By Taylor's formula we have

$$\begin{aligned}\sqrt{nh}\,(\hat{g}_{\tau+1}(\xi;\hat{\alpha})-\mathrm{E}(\hat{g}_{\tau+1}(\xi;\alpha))) &= \sqrt{nh}\,(\hat{g}_{\tau+1}(\xi;\alpha)+\hat{g}'_{\tau+1}(\xi;\alpha_*)(\hat{\alpha}-\alpha)-\mathrm{E}(\hat{g}_{\tau+1}(\xi;\alpha)))\\ &= \sqrt{nh}\,(\hat{g}_{\tau+1}(\xi;\alpha)-\mathrm{E}(\hat{g}_{\tau+1}(\xi;\alpha)))+\sqrt{nh}\,\hat{g}'_{\tau+1}(\xi;\alpha_*)(\hat{\alpha}-\alpha)\end{aligned}$$

where α_* lies between $\hat{\alpha}$ and α. Since $\hat{\alpha} = \alpha + o_p(1)$ in line with Theorem 3, it suffices to prove that $\sqrt{nh}\,\hat{g}'_{\tau+1}(\xi;\alpha_*) = O_p(1)$. We have (with $\omega'_t \equiv \mathrm{d}\omega(t)/\mathrm{d}\alpha$)

$$\mathrm{E}[\hat{g}'_{\tau+1}(\xi;\alpha)] = \mathrm{E}\left(\sum_{t=1}^{\tau}\hat{f}_t(\xi)\omega'_s\right) = \sum_{t=1}^{\tau} f_t(\xi)\omega'_s + O(h^2)\sum_{t=1}^{\tau}\omega'_s$$

$$\begin{aligned}\mathrm{var}(g'_{\tau+1}(\xi;\alpha)) &= \mathrm{var}\left[\sum_{t=1}^{\tau}\hat{f}_t(\xi)\omega'_t\right]\\ &= \sum_{t=1}^{\tau}\mathrm{var}\left[\hat{f}_t(\xi)\omega'_t\right]+2\sum_{t=1}^{\tau}\sum_{s=1}^{t-1}\mathrm{cov}\left(\hat{f}_t(\xi),\hat{f}_s(\xi)\right)\omega'_t\omega'_s\\ &= \sum_{t=1}^{\tau}(\omega'_t)^2\,\mathrm{var}\left[\hat{f}_t(\xi)\right]+\\ &\quad +2\sum_{t=1}^{\tau}\sum_{s=1}^{t-1}\frac{1}{nh}\rho_{(t-s)}\left(\int K^2(u)\,f_t(hu+\xi)\,\mathrm{d}u-O(h)\right)\omega'_t\omega'_s\\ &= O\left(\frac{1}{nh}\right)\sum_{t=1}^{\tau}(\omega'_t)^2+O\left(\frac{1}{nh}\right)\sum_{t=1}^{\tau}\sum_{s=1}^{t-1}\rho_{(t-s)}\omega'_t\omega'_s\end{aligned}$$

Note that $\sum_{t=1}^{\tau}(\omega'_t)^2 = \frac{2(-\alpha+\alpha^{1+2\tau}-\alpha^{\tau}\tau(\alpha^2))}{(\alpha-1)\alpha(1+\alpha)^3(\alpha^{\tau}-1)^2} \to 2/(1-\alpha)(1+\alpha)^2$ is convergent. The term $\sum_{t=1}^{\tau}\sum_{s=1}^{t-1}\rho_{(t-s)}\omega'_t\omega'_s$ is obviously bounded for $\tau < \infty$. Even when $\tau \to \infty$, the sum is finite. To confirm this, let $\rho_{(t-s)} = 1$, $\forall t,s$. We have $\left|\sum_{t=1}^{\tau}\sum_{s=1}^{t-1}\rho_{(t-s)}\omega'_t\omega'_s\right| < \sum_{t=1}^{\tau}\sum_{s=1}^{t-1}|\omega'_t||\omega'_s| = \sum_{t=1}^{\tau}|\omega'_t|\sum_{s=1}^{t-1}|\omega'_s| \le \sum_{t=1}^{\tau}(\omega'_t)^2\sum_{s=1}^{t-1}(\omega'_s)^2$ and this series

DOI: 10.1002/for

is convergent, as shown previously. The conclusion is that $\mathrm{var}(\hat{g}'_{\tau+1}(\xi;\alpha)) = O\left((nh)^{-1}\right)$, which implies that $\sqrt{nh}\,\hat{g}'_{\tau+1}(\xi;\alpha_*) = O_p(1)$. □

ACKNOWLEDGEMENTS

I am grateful to an anonymous referee for very helpful comments and suggestions. This research was supported by the Fundação para a Ciência e a Tecnologia (programa FEDER/POCI 2010).

REFERENCES

Davidson J. 1994. *Stochastic Limit Theory*. Oxford University Press: Oxford.

Gouriéroux C, Monfort A. 1995. *Statistics and Econometric Models*, Vol. II. Cambridge University Press: Cambridge, UK.

Pagan A, Ullah A. 1999. *Nonparametric Econometrics*. Cambridge University Press: Cambridge, UK.

Pesaran M. 2007. A simple panel unit root test in the presence of cross section dependence. *Journal of Applied Econometrics* **22**(2): 265–312.

Author's biography:
João Nicolau is working as an Associate Professor with '*Agregação*' at Technical University of Lisbon/School of Economics and Management (ISEG). He is the head of the Econometrics and Time Series Group of CEMAPRE (Centre for Applied Mathematics and Economics) and the Econometrics Area at the Mathematics Department of ISEG. He is also the coordinator of the Academic Master's Program in Applied Econometrics and Forecasting at ISEG.

Author's address:
João Nicolau, School of Economics and Management (ISEG)/Technical University of Lisbon (UTL) and CEMAPRE, ISEG, Rua do Quelhas 6, 1200-781 Lisbon, Portugal.