ETLA ELINKEINOELÄMÄN TUTKIMUSLAITOS
THE RESEARCH INSTITUTE OF THE FINNISH ECONOMY

ETLA Raportit
ETLA Reports

25.8.2014

No 24

Nonparametric Estimation of Conditional Expectations for Sustainability Analyses

Juha Alho*

* University of Helsinki, Social Statistics, juha.alho@helsinki.fi

Suositeltava lähdeviittaus: Alho, Juha (25.8.2014). "Nonparametric Estimation of Conditional Expectations for Sustainability Analyses". ETLA Raportit No 24. http://pub.etla.fi/ETLA-Raportit-Reports-24.pdf

ISSN-L 2323-2447
ISSN 2323-2447 (print)
ISSN 2323-2455 (online)

Sisällysluettelo

Nonparametric Estimation of Conditional Expectations for Sustainability Analyses

Abstract
Optimal forecasts are, under a squared error loss, conditional expectations of the unknown future values of interest. When stochastic demographic models are used in macroeconomic analyses, it becomes important to be able to handle updated forecasts. That is, when population development turns out to differ from the expected one, the decision makers in the macroeconomic models may change their behavior. To allow for this, numerical methods have been developed that allow us to approximate how future forecasts might look like, for any given observed path. Some technical details of how this can be done in the R environment are given.

Key words: Demography, forecasting, overlapping generations

Kestävyysanalyyseissä tarvittavien ehdollisten odotusarvojen laskenta epäparametrisillä menetelmillä

Tiivistelmä
Pienintä neliövirhettä vastaavat optimaaliset ennusteet ovat odotusarvoja. Kun stokastisia demografisia malleja käytetään makroekonomisissa kestävyystarkasteluissa, on tärkeää voida käsitellä sellaisia tilanteita, joissa päätöksentekijät muuttavat käytöstään uuden informaation tullessa tietoon. Käytännössä tämä tarkoittaa päivittyvien ennusteiden laskentaa. Kirjoituksessa käsitellään lähestymistapoja, joita voidaan R-ympäristössä käyttää tarvittavien ehdollisten odotusarvojen laskennassa.

Asiasanat: Väestötiede, ennustaminen, sukupolvimalli

1 Introduction

In the analysis of fiscal sustainability, decisions concerning current and future labor demand and supply depend on population forecasts, as they tell us what the future potential labor supply in different ages will be. When population development deviates from the forecast, the decision makers might like to adjust their decisions using an updated forecast. The calculations described in this paper are needed in the modeling of the behavior of an economy in which such adjustments do occur.

This topic is closely related to the literature on forecast updating in time-series analysis (e.g., Box and Jenkins (1976), p. 164). However, those discussions typically assume that a well-identified model is available. Then, the required conditional expectations of future values can be derived either analytically, or by iterative numerical methods.

In our applications this is not the case. Instead, a recursive equation for the renewal of the population vector is used repeatedly based on simulated values of the future vital rates (of fertility, mortality, and migration). This produces a numerical approximation for the predictive distribution of the future population vector for lead times of interest. The question we ask in this paper is, how can one best use the simulated counts to estimate the future expected values of the population counts, conditionally on how the population has evolved earlier? A more detailed exposition of the relevance of the setting is given in Alho (2014).

In an Appendix we give some R scripts that can be used to implement the proposed procedures. In the practical applications we will consider, the program PEP (Program for Error Propagation) has been used to generate the random population counts.

2 Cohorts as Overlapping Generations

In the so-called *overlapping generations* (OLG) model decision makers consider the consumption and saving behavior of all current and future birth cohorts. For example, in the Finnish Overlapping Generations (FOG) model of Lassila, Palm and Valkonen (1997), there is a profit-maximizing firm that pays wages to labor force. In return, the labor force agrees to work to finance consumption. The price of labor depends on the size of working age cohorts. Thus, the optimal labor supply, wages, and consumption decisions of all currently living and future cohorts are interlinked. In the traditional approach, the future population evolution is assumed to follow the best estimate path, and the wages, labor supply, and consumption are optimized for all current

and future cohorts in a single step.

In Alho et al. (2008) several applications of OLG models are given, in which stochastic population evolution is allowed. This is accomplished by repeatedly generating sample paths representing alternative future developments of the population. If the OLG model is solved path by path, this would be equivalent to assuming that the decision makers possess the gift of *perfect foresight*. A more realistic approach is to assume that the decision makers revise their decisions as it becomes obvious that the population has evolved differently from the expected path (cf. Alho and Määttänen (2008)).

3 Linear Growth Model for Population

Let $\mathbf{V}(t) = (\mathbf{V}(0,t)^T, \ldots, \mathbf{V}(\omega,t)^T)^T$ represent the population in the beginning of year t, where the vectors $\mathbf{V}(x,t)$ have the numbers of females and males in ages $x = 0, \ldots, \omega$. The population one year later can be written in the form

$$\mathbf{V}(t+1) = \mathbf{R}(t)\mathbf{V}(t) + \mathbf{N}(t), \tag{1}$$

where $\mathbf{R}(t) = (\mathbf{R}(x,y,t))$ is a matrix of $(\omega+1) \times (\omega+1)$ blocks, and $\mathbf{N}(t)$ is a vector of the same form as $\mathbf{V}(t)$ that contains the net number of migrants surviving to the beginning of year $t+1$. The first column of the matrices $\mathbf{R}(x,0,t)$ contains the expected numbers of girls and boys born per woman in child bearing ages, say, $x = 15, \ldots, 49$. The second column is zero. The subdiagonal blocks $\mathbf{R}(x+1,x,t)$ contain survival probabilities from age x to age $x+1$, and $\mathbf{R}(\omega,\omega,t)$ contains probabilities of survival in the open-ended highest age, for example $\omega = 100+$. Other blocks are zero (Alho and Spencer (2005), pp. 180-183).

When population renewal is stochastic, the matrices $\mathbf{R}(t)$ and the vectors $\mathbf{N}(t)$ are random.

4 Nonparametric Estimation of Conditional Expectations

4.1 Using a Neighborhood Average

Let $\mathbf{Y}(t,H)$ denote a vector of all population counts during years $t, \ldots, H$ (in some order). The predictive distribution of such vectors is assumed to be continuous but its exact form, while fully specified by (1), is, in practice, only known up to the accuracy of simulations. The probability of any sample path is zero, or $P(\mathbf{Y}(0,T) = \mathbf{y}(0,T)) = 0$ for any $\mathbf{y}(0,T)$. Therefore, the

conditional expectation $E[\mathbf{Y}(T+1,H) \mid \mathbf{Y}(0,T) = \mathbf{y}(0,T)]$ cannot be estimated based on the simulations, as in any finite (or even countably infinite) set of simulations, there is, with probability one, only one path that has the beginning $\mathbf{y}(0,T)$.

A way out can be based on the approximation

$$\begin{aligned} &E[\mathbf{Y}(T+1,H) \mid \mathbf{Y}(0,T) = \mathbf{y}(0,T)] \\ \approx\ &E[\mathbf{Y}(T+1,H) \mid \mathbf{Y}(0,T) \in U(\mathbf{y}(0,T),c)], \end{aligned} \tag{2}$$

where $U(\mathbf{y}(0,T),c) = \{\mathbf{y} \mid d(\mathbf{y} - \mathbf{y}(0,T)) < c\}$ and d is some metric. The approximation becomes an identity as $c \downarrow 0$, but for any $c > 0$ the estimate is typically biased.

Our proposed approach is based on the following. Suppose we have N samples taken from the predictive distribution of $\mathbf{Y}(1,L)$ that we wish to use for a sustainability analysis. We will then take $N' >> N$ supplementary samples from the same distribution.[1] These will be used to compute the neighborhood averages.

To develop some intuition, consider scalar-valued Y_1 and Y_2. Suppose that these are related linearly $Y_2 = a + bY_1 + \epsilon$, where $E[\epsilon \mid Y_1] = 0$ and $Var(\epsilon \mid Y_1) = \sigma^2$. Pick some $c > 0$ and a value y_1, and write $P(Y_1 \in U(y_1,c)) = \delta$, where $d(y - y_1) \equiv \mid y - y_1 \mid < c$. Suppose we have a sample of size N' available from the distribution of (Y_1, Y_2). Then, the bias of the estimate is

$$E[Y_2 \mid Y_1 \in U(y_1,c)] - E[Y_2 \mid Y_1 = y_1] = b(E[Y_1 \mid Y_1 \in U(y_1,c)] - y_1). \tag{3}$$

This depends on b and the bias in the sampling of Y_1 values. To see how this works, suppose that the density is linear in the neighborhood, $f(y) = \kappa_0 + \kappa_1(y - y_1)$, so that $f(y_1) = \kappa_0$ and $f'(y_1) = \kappa_1$, and where $\kappa_0 \geq c \mid \kappa_1 \mid$. Then, setting $P(Y_1 \in [y_1 - c, y_1 + c]) = \delta$, we get that $2\kappa_0 c = \delta$, and $E[Y_1 \mid Y_1 \in [y_1 - c, y_1 + c]] = y_1 + \kappa_1 c^2/3\kappa_0$. This shows that the bias vanishes as $c \downarrow 0$, but we also see that the steeper the tangent line, the larger the bias.

The mean squared error of the estimator of the conditional expectation has also a variance component. Suppose for simplicity that the error term ϵ is independent of Y_1. If the number of sample points falling in the neighborhood is n, then the variance of the sample mean is σ^2/n. A first order approximation for the unconditional variance is $\sigma^2/E[n]$. This shows that if one wishes to reduce the bias by taking c small, one has to pay for that in terms of small expected sample $E[n] = N'\delta$ and correspondingly larger random error.

[1] In practice, this is accomplished by running the recursion (1) a total of $N + N'$ times for different sets of demographic rates.

Continuing with the assumption that the density is linear in the neighborhood, note that δ depends on y_1 and c. Suppose we fix the value of δ first, and determine c so that $P(Y_1 \in [y_1 - c, y_1 + c]) = \delta$ is true for all values of y_1 and c. Under the linear case, this is accomplished by taking $c = \delta/2\kappa_0$. This would guarantee that around each target value of interest y_1 the expected number of samples would be $E[n] = \delta N'$, where N' is the number of supplementary samples. A simple alternative is that one fixes $n = \delta N'$, and lets c vary freely.

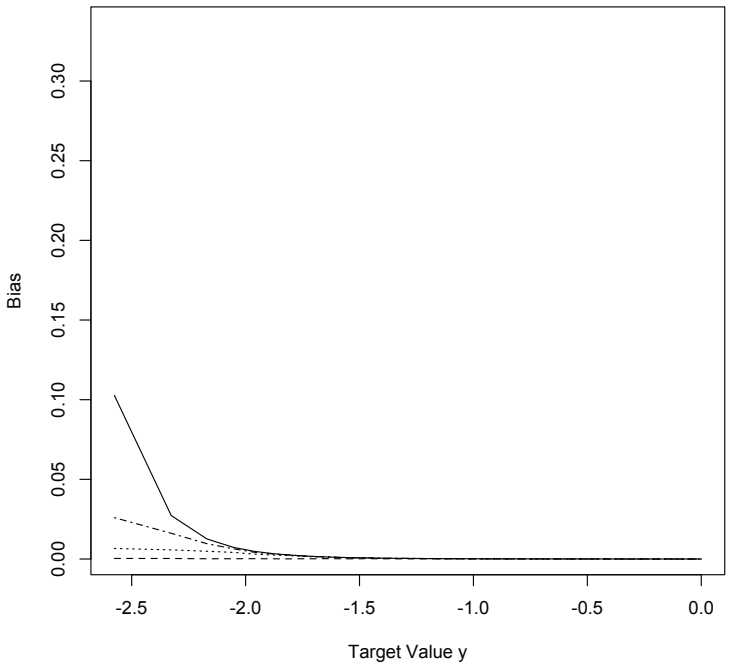

Figure 1: Bias of the Kernel Estimator at Quantiles of the Standard Normal Distribution (Target Values) Corresponding to Probabilities $0.005, 0.010, \ldots, 0.500$ for $h = 0.02$ (dashed), $h = 0.05$ (dotted), $h = 0.10$ (dash-dotted) and $h = +\infty$ (solid).

Example 1. In Figure 1, we assume that $f(y) = \exp(-y^2/2)/\sqrt{2\pi}$, i.e., it is the density of the standard normal distribution. We consider the mean of the sampled y values, around y_1 values that correspond to quantiles for probabilities $= 0.005, 0.010, \ldots, 0.500$ (the results for larger quantiles are symmetric). We take $N' = 10,000, \delta = 0.01$ and $c = \delta/2\kappa_0$, so each neighborhood $U(y_1, c)$ is expected to have 100 supplementary samples. The values

of y_1 increase from -2.58 to 0, and correspondingly the neighborhood radiuses decrease from 0.346 to 0.0125, first rapidly, and then leveling off. The bias using a local linear approximation to the normal density is given by the solid line. We see that up to two standard deviations, the theoretical bias is $<$ 0.0075. While the bias seems small, less than 0.5 % of the target value y_1 that we find on the horizontal axis, one can reduce the bias further.

4.2 Nearest Neighbor Approach

An alternative to using an average over a neighborhood are the nearest neighbor estimators that use some form of weighted averaging. Let us write $E[\mathbf{Y}_2 \mid \mathbf{Y}_1 = \mathbf{y}_1] = g(\mathbf{y}_1)$, where we can identify $\mathbf{Y}_1 = \mathbf{Y}(0,T)$ and $\mathbf{Y}_2 = \mathbf{Y}(T+1,h)$. Then we have the model

$$\mathbf{Y}_2 = g(\mathbf{y}_1) + \epsilon(\mathbf{y}_1), \tag{4}$$

where $E[\epsilon(\mathbf{y_1})] = \mathbf{0}$. The function g is smooth, because the population renewal model (1) has derivatives of all orders. Suppose that we wish to estimate the conditional expectation $g(\mathbf{y}_1)$ and we have the sample $(\mathbf{y}_{1i}, \mathbf{y}_{2i}), i = 1, 2, \ldots, n$ from the joint distribution of $(\mathbf{Y}_1, \mathbf{Y}_2)$. An *n-nearest neighbor estimator* would be

$$\hat{g}(\mathbf{y}_1) = \sum_{i=1}^{n} \mathbf{y}_{2i} K(\mathbf{y}_1 - \mathbf{y}_{1i}) / \sum_{i=1}^{n} K(\mathbf{y}_1 - \mathbf{y}_{1i}), \tag{5}$$

where the *kernel* K is some unimodal probability density symmetric around the origin, and the sum is extended over those sample points with the n smallest distances $d(\mathbf{y}_{1i} - \mathbf{y}_1)$. To illustrate this approach, we return to the example of the previous section.

Example 2. Continuing with Example 1, we note that the most widely used kernel is the Gaussian, with $K(x) = exp(-x^2/2h^2)/\sqrt{2\pi}h$ (cf., Silverman (1998)). Then, the bias caused by having to use of a biased sample of y_{1i}'s equals

$$B(y_1) = \sum_{i=1}^{n} E[Y_{2i} K(y_1 - Y_{1i}) / \sum_{i=1}^{n} K(y_1 - Y_{1i})] - y_1. \tag{6}$$

As above, this effect is further multiplied by b. If we choose the bandwidth roughly as $h \sim c$, then the nonparameteric estimate would be approximately a weighted average of the points in the neighborhood $U(y_1, c)$. The radiuses c range in this example approximately from 0.0125 to 0.346. Recall that as we use a standard normal distribution as our model of Y_1's, the radiuses

can be interpreted as going from 1.25 % to 34.6 % of the standard deviation with the largest value at the tails. In Figure 1, we illustrate the bias for $h = 0.02, 0.05, 0.10$. The finding is clear, the smaller h is, the smaller is the bias.

4.3 Bias-Variance Trade-off

The nearest neighbor estimator is biased, but the bias is expected to be small, except at extreme tails of the density that have a very small probability. Moreover, as shown by Figure 1, even at the tails the bias can be made small by selecting small values for the scale parameter. The downside is that the *effective sample size* is reduced. To get an idea of what this could mean in practice, let us write for the (random) weights in formula (5) as $w_i \equiv K(y_1 - Y_{1i}) / \sum_{i=1}^{n} K(y_1 - Y_{1i})$, for short. If we condition on the observed values Y_{1i}, and assume that the error terms ϵ are independent of them, with $Var(\epsilon_i) = \sigma^2$, then, the variance of the kernel estimate is

$$\sigma^2 \sum_{i=1}^{n} w_i^2. \tag{7}$$

In case the weights are all equal, or $w_i \equiv 1/n$, then the variance is simply the usual variance of the sample mean σ^2/n. However, when the weights are not constant, the variance increases.

Example 3. Using the Gaussian kernel of Example 2, we find that the expected value of the sum in (6) is,

$$I(h) \equiv \sum_{i=1}^{n} (e^{-z_i^2/2h^2})^2 / (\sum_{i=1}^{n} e^{-z_i^2/2h^2})^2, \tag{8}$$

where the summation is over the $n = 100$ nearest neighbors of y_1. Complementing the bias results of Examples 1 and 2, we can now calculate the inflation factor $I(h)$ for $h = 0.02, 0.05, 0.10$. In Figure 2 these values are compared to $I(+\infty) = 1/n = 0.01$ that corresponds to simple random sampling. The interpretation is that e.g. for $h = 0.02$ and $y_1 = -2.0$ the sampling variance of the kernel estimator is approximately 5 times as large as that of the unweighted neighborhood average. As expected, we see the reverse order, as compared to Figure 1. The small values of h that are so successful in reducing the bias, now lead to inordinate increases in sampling variability. For comparison with the results of the previous examples, we may set $\sigma = 1$. This corresponds to a random walk model where $a = 0, b = 1$, and Y_1 is the first observation and Y_2 is the second. It seems that qualitatively the larger

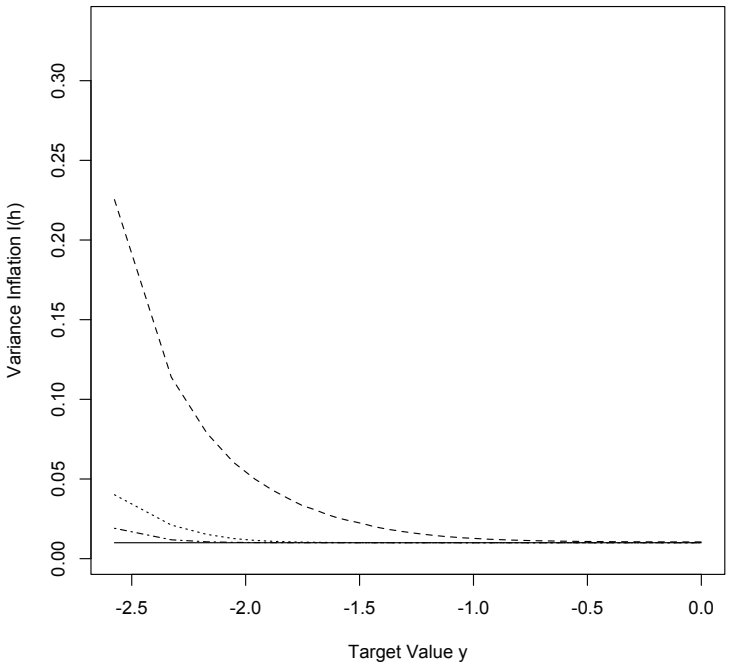

Figure 2: Variance Inflation Factor of the Kernel Estimator at Quantiles of the Standard Normal Distribution (Target Values) Corresponding to Probabilities $0.005, 0.010, \ldots, 0.500$ for $h = 0.02$ (dashed), $h = 0.05$ (dotted), $h = 0.10$ (dash-dotted) and $h = +\infty$ (solid).

issue is the variance inflation rather than the bias. In fact, in the computation of the *root mean squared error* the bias estimates of Figure 1 will be squared before adding to the variances of Figure 2, so they are an order of magnitude smaller. This is confirmed by Figure 3 that gives the root mean squared errors for different target values. In any case, a compromise between bias and sampling variability may be sought. For these particular data the optimal weights would be in the neighborhood $0.05 < h < 0.10$.

4.4 Repeated Conditional Expectations

The previous example already alluded to the aspect that we have so far ignored. That is, the updating of the conditional expectation is carried out repeatedly. Here, two aspects seem relevant. First, several aspects of the recursive process of population renewal are non-Markovian, and we would like to account for this possibility. The second aspect is the non-stationary

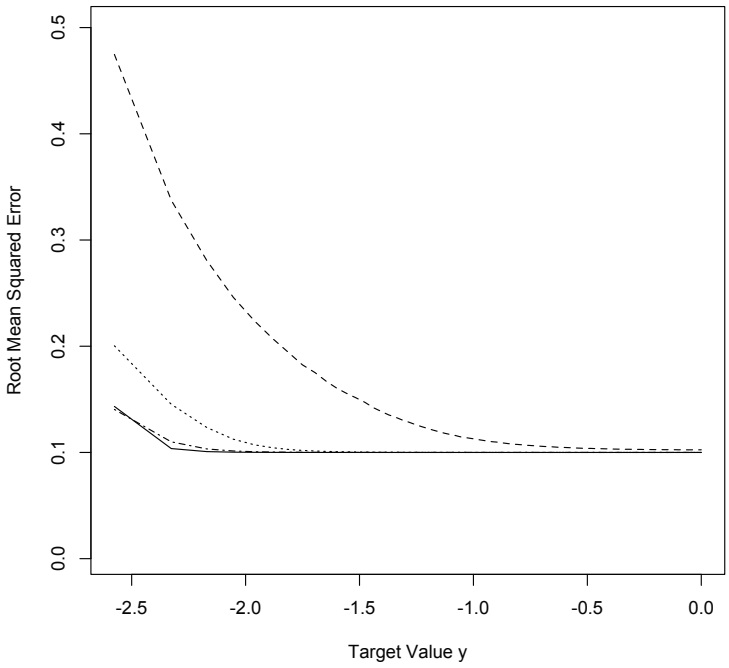

Figure 3: Root Mean Squared Error of the Kernel Estimator at Quantiles of the Standard Normal Distribution (Target Values) Corresponding to Probabilities $0.005, 0.010, \ldots, 0.500$ for $h = 0.02$ (dashed), $h = 0.05$ (dotted), $h = 0.10$ (dash-dotted) and $h = +\infty$ (solid).

nature of the renewal process. In non-stationary processes we would typically see some sample paths that would systematically deviate from the expected. For such sample paths, forecast errors would be expected to be large, but our computational methods should not exaggerate the errors unnecessarily. Therefore, we would like to make our procedures robust against errors incurring for such paths.

To study this, we will expand the setting of Examples 1 and 2 to include several updates. Suppose $X_i \sim N(0,1), i = 1, 2, \ldots, L$ are independent, and define first $Y_1 = X_1$, and then recursively $Y_t = Y_{t-1} + X_t$ for $t = 2, \ldots, L$. In other words, the variables Y_t are observations of a standard Brownian motion at $t = 1, 2, \ldots, L$. It follows that $Y_t \sim N(0, t)$ so that the best forecast is initially $\hat{Y}_t = 0$. A particularly difficult case for the nonparametric estimation is that the target sample path deviates increasingly from 0, for

example $y_t = B\sqrt{t}$ for some constant B. [2] Under the Brownian motion model the best forecast for future values Y_{t+k} at time t would then be the current value $B\sqrt{t}$. In this case the true underlying process is normal and Markovian, but in our real applications neither property would hold exactly, so potentially the whole history up to time t might be informative.

One way to take this into account is by by taking, at time $t = 1$ a larger sample, say of size N_1', consisting of those sample paths whose values at $t = 1$ are closest to y_1, and then at time $t = 2$ leaving out some sample paths that are furthest from y_2 etc. In this manner we would have a nested sequence of subsets of sizes $N_1' > N_2' > \ldots > N_{L-1}'$. Suppose that at each time t we would use the neighborhood average to compute the conditional expectations of future values based on the subset then remaining. If the target sample path would have a very high likelihood in the beginning, the gradual depletion of the sample would be wasteful, but on the other hand, if the target sample path would be of the type envisioned in the previous paragraph, the influence of far away points need to be reduced. Yet, some simple calculations suggest that simply depleting the neighborhoods is not the preferred way to go.

Consider the path $y_t = B\sqrt{t}$, for example. Let the first value be $y_1 = B\sqrt{1} = B$. Choose $0 < \delta < 1$ and out of the supplementary samples select those $\delta N'$ sample paths whose value at $t = 1$ is closest to B. Their average would have some bias, as discussed in Examples 1 and 2, but this would be small. However, at time $t = 2$ the value of the target process would be $B\sqrt{2}$ and at this point the average of the original sample for future times $t = 3, \ldots, L$ would continue to be B. This could not be considered as an acceptable numerical approximation for the conditional expectation as it would ignore the knowledge that the process is currently at $B\sqrt{2}$. Suppose for definiteness that $B < 0$. Then, one would wish to remove some of the sample paths with the largest values at $t = 2$, so as to correct for the bias.

More generally, consider any $t = 2, \ldots, L-1$. To reduce bias, with $B < 0$ we would want to eliminate all values beyond a threshold at Q_s such that

$$\int_{-\infty}^{Q_s} \frac{z}{\sqrt{2\pi}} e^{-z^2/2} dz = B\sqrt{s} \int_{-\infty}^{Q_s} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz \tag{9}$$

for $s = 1, \ldots, t$. Under this procedure the conditional expectation of the remaining sample would equal the target value $B\sqrt{t}$. Such threshold values obviously depend on B. The values of Q_t are a bit difficult to assess directly. However, we can translate them into probability terms by giving the fraction of the original data that are to be excluded at each time t. Equivalently, we

[2] This choice is not a "typical" Brownian motion path, but it is chosen here for numerical convenience.

consider the fraction remaining out of the original sample. The analytical evaluation of (9) becomes messy as the process is iterated for $s = 1, \ldots, t$, but simulation can be used. Based on 100,000 simulated random walks we estimate that at $t = 1$ the fraction remaining is approximately 0.70, 0.38, 0.17, 0.06, and 0.02 for $B = 0.5, 1.0, 1.5, 2.0, 2.5$. Complementing these estimates Figure 4 shows how the fraction remaining in the sample, *as compared to the sample remaining at* $t = 1$ declines for $t = 1, 2 \ldots, 50$. We see that the patterns of decline is very similar, but as the sample sizes themselves depend radically on B at $t = 1$ already, the actual remaining sample sizes have different orders of magnitude.

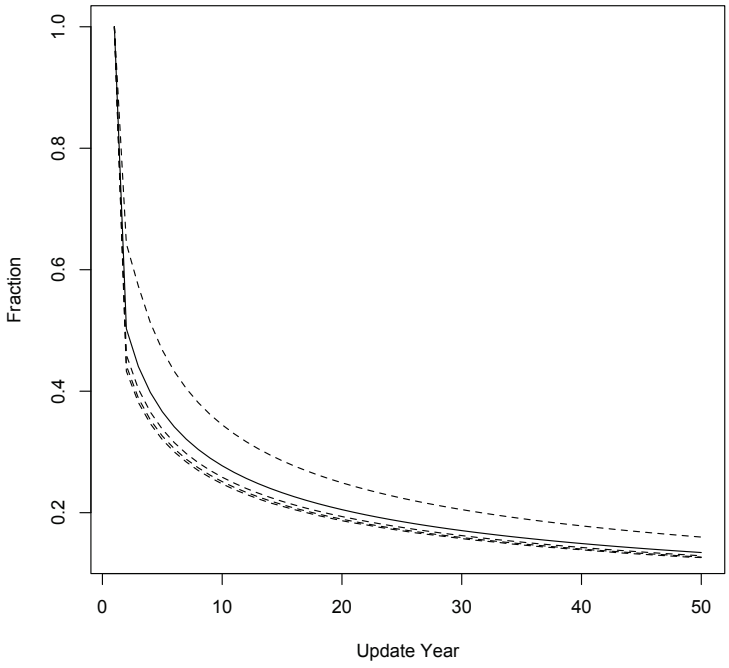

Figure 4: Fraction of the Sample Remaining As a Function of Update Year for Path $y_t = B\sqrt{t}$, from Top to Bottom Corresponding to $B = 0.5, 1.0, 1.5, 2.0, 2.5$; the Graph for $B = 1.0$ is Solid.

The general finding is that the requirement of unbiasedness can lead to a very rapidly depleting sample size for paths that veer systematically off the target. This suggests that instead of reducing the size of the neighborhood, it is preferable to keep the neighborhood that is initially selected, but use changing weights for different lead times to reduce the bias. At the same

time, we have to watch out for schemes that lead to a too rapidly declining *effective* sample sizes.

4.5 Practical Implementation of Non-parametric Estimation for Population Renewal Problems

All our illustrations have been made in the scalar valued setting, to see what happens when the details are fully known. However, in our applications the setting is potentially a very high-dimensional one. If population is given by single years of age (say, $x = 0, 1, \ldots, 100$), and there are H future years of interest, then the vector of future values for which a conditional expectation is desired has dimension $2 \times 101 \times H$. The components of the vector are statistically dependent via the renewal equations and the the correlatedness of the mortality and fertility rates and the net-migration numbers.

4.6 Forecasting Increments

A practical aspect is the starting value of the updated conditional expectation calculations. Continuing with the notation of the previous discussion, suppose the population at the time of the update is denoted by $\mathbf{y}_1$, the simulated values for the population in the (χ^2-distance) neighborhood are $\mathbf{Y}_{1i}, i, \ldots, n$, and suppose the simulated population counts for some later year are noted as $\mathbf{Y}_{2i}, i = 1, \ldots, n$. Then, it seems natural to base the conditional expectation on the latest observed value, and use the supplementary paths to provide the increments, so the updated forecast for the later time point would be

$$\hat{\mathbf{Y}}_2 = \mathbf{y}_1 + \frac{1}{n}\sum_{i=1}^{n}(\mathbf{Y}_{2i} - \mathbf{Y}_{1i}) \tag{10}$$

This has the advantage that it *eliminates the jump-off error.*

4.7 Choosing a Metric for Weighting

Based on the considerations of Section 4.4 we will keep the neighborhood as fixed, after it has first been chosen, but allow weights to adapt within the neighborhood, as more is learned with increasing lead times. After some experimentation the following heuristic approach was developed for the n-nearest neighbor estimation.

First, distance between a given target count $\mathbf{V}(t)$ and the corresponding supplementary samples $\mathbf{V}_i(t), i = 1, \ldots, N'$ we have used the so-called χ^2-distance

$$D_i(t) \equiv (\mathbf{V}(t) - \mathbf{V}_i(t))^T D(\mathbf{V}(t))^{-1}(\mathbf{V}(t) - \mathbf{V}_i(t)), \tag{11}$$

where $D(\mathbf{V}(t))$ is a diagonal matrix with diagonal elements $\mathbf{V}(t)$. This measure controls relative error. [3]

Second, at first update year $T = 1$ a neighborhood of size n of the given sample path was chosen with the smallest distances $D_i(1)$. This subset of supplementary sample paths was kept fixed in forecast updates at $T > 1$, but, reweighted each year. For convenience, we index the neighborhood paths by $i = 1, \ldots, n$.

Third, given the rapid increase of variability, the $D_i(t)$ tended to increase rapidly with t. To make the values comparable for different t, we computed each year $med(t)$ = the median of $D_i(t), i = 1, \ldots, n$, and used $D_i(t)/med(t)$ as the measure of distance. The reweighting involved two parameters $u > 0$ and $0 < q \leq 1$ so that the kernel function evaluated for sample path i during update year T was $K_i(T) \equiv exp(-uq^{1-T}D_i(T)/med(T))$. The intuition is that u gives the average level of decay for the weights, and q determines how rapidly the decay should increase with T.

Under this formulation, the nonparametric regressions depended on the parameters $\theta \equiv (n, u, q)$, where n/N' can be interpreted as the δ of Section 4.1. Alho (2014) discusses further, how *cross validation* can be used to select the parameters. Very briefly, a systematic grid search can be performed to find a parameter combination that reduces the distance between n-nearest neighbor estimator and the (known) target value. This proved to be quite time consuming.

We have written short scripts in R (http://CRAN.R-project.org) that provide a simple illustration of how the preceding approach can be numerically implemented. This is described in the Appendix.

References

Alho J.M. (1990) Stochastic methods in population forecasting. *International Journal of Forecasting*, 6, 521-530.

Alho J.M. (2014) Forecasting Demographic Forecasts. *International Journal of Forecasting*, to appear.

Alho J.M., Hougaard Jensen S.E. and Lassila J. (Eds.)(2008) *Uncertain Demographics and Fiscal Sustainability*. Cambridge: Cambridge University Press.

[3] On the other hand, the very oldest age groups (95+) could be excluded from the distance calculation, since they do not participate in the labor market and their total pension cost is small.

Alho J. and Keilman N. (2010) On future household structure. *Journal of the Royal Statistical Society*, A,173, 117-143.

Alho J.M. and Määttänen N. (2008) Informational assumptions, aggregate mortality risk and life-cycle saving, pp. 219-238 in Alho J.M., Hougaard Jensen S.E. and Lassila J. (Eds.)(2008).

Alho J.M. and Spencer B.D. (2005) *Statistical Demography and Forecasting*. Nerw York: Springer.

Box G.E.P. and Jenkins G.M. (1976) *Time-series analysis, forecasting and control, 2nd ed.* SanFrancisco: Holden Day.

Lassila J., Palm H. and Valkonen T. (1997) FOG: The Finnish Overlapping Generations model. ETLA, Discussion papers 601. Helsinki.

Silverman, B.W. (1998) *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall/CRC.

A Appendix. Directions for the Use of FPATH Commands Given as R Scripts with PEP output

R-version of FPATH programs uses files created by programs PEP and COMBINE as input. It produces numerical estimates of conditional forecasts for future population for target sample paths. These are given as text files.

To use the FPATH programs, do the following. (The directions are tailored for use in Windows.)

0. Create a folder D:\fin09 (change the name to whatever you wish to use).

1. Produce a forecast for Finland with jump-off time January 1, 2009, for years 2010-2060 (i.e. 51 forecast years), using PEP, with output in the folder D:\fin09. After the run, aggregated files for ages 0-4, 5-9, ..., 105+, both sexes combined, are assumed to exist. I.e., there are 22 age-groups, and the number of sample paths is 9,300. The first 300 are the target paths, and the remaining 9,000 are used to form the neighborhood samples. File names are of the form D:\fin09\Pxxxx_0.C1, D:\fin09\Yyy_0.C1.

2. Start R. Copy the text in file fpath-input.txt, and paste it into R. This file is given below. This sets up the necessary input data, but does not

calculate anything. Save the workspace (from the pull-down menu File of R) into file D:\fin09\f09.RData. (Exit R at this point if you like.)

3. (If you exited, start R again by clicking on the icon of file D:\fin09\f09.RI Copy the text in file fpath-output.txt, and paste it into R. This produces FPATH output in the form of text files D:\fin09\Fzzz_0.DAT. There will be 300 such files.

4. The files D:\fin09\Fyyy_0.DAT have 22 columns, one for each age group. There are a total of 55 lines. These form 10 blocks: the first has 10 lines, the second 9 lines,..., the tenth has 1 line.

The first block differs from the rest. The first line has the point forecast for 2014, the second for 2019,..., the tenth has the point forecast for 2059. These are taken directly from the file D:\fin09\P0_0.C1, lines 5, 10, 15,...,50, and they are identical for all files D:\fin09\Fzzz_0.DAT. The remaining lines differ according to sample path to which the file corresponds.

The first line of second block gives the conditional forecast for year 2019 given that data of 2014 have been observed, the second line of second block gives the conditional forecast for 2024 given that data of 2014 have been observed,..., the ninth line of second block gives the conditional forecast of 2059 given that data of 2014 have been observed. These, and subsequent values are computed from files D:\fin09\Yyy_0.C1.

The eight lines of third block give conditional forecasts for 2024, 2029,..., 2059, given that data of 2019 have been observed.

Similarly, for the other blocks, so that the one line of tenth block gives a conditional forecast for 2059 given that data of 2054 have been observed.

5. The intended use of the FPATH output files. The Decision Makers of the OLG model are assumed, first to make a lifetime plan in the beginning of 2009, based on the point forecast that is given by the ten lines of the first block, for 2014-2059, for each of the three hundred sample paths. These plans are identical. After the population evolution until the beginning of 2014 becomes known, the Decision Makers will change their plan according to what now seems optimal, in view of the conditional forecast given in the second block for 2019-2059.

6. FPATH calculations have not been extensively tested, nor have the calculations of the conditional expectations been optimized. The latter involves a study of how the subset of paths is best chosen for these calculations, for each path. In this example, 350 paths are chosen at first step based on a Chi-squared distance. The nearest neighbor kernel estimator involves two parameters that regulate the weights at different lead times.

A.1 fpath-input

```
#fpath-input.txt
#compute conditional expectations from PEP-output
#n = number of output paths and conditional forecast files
#N = number of files available for the computation of conditional expectations
#S = pick every Sth year
#J = number of age-groups
#ne = number of vector paths that form the neighborhood of the output path
#fy = number of forecast years in the source forecast produced with PEP
n<-300
N<-9000
S<-5
T<-10
J<-22
ne<-100
fy<-51
#a is used for the computation of distances
a<-as.vector(rep(1,N))
#end and beginning rows of blocks in output files
e<-cumsum(T:1)
st<-c(0,e[c(1:T-1)])+1
#addresses of files for input and output
os1<-"D:\\fin09\\P0_0.C1"
os2<-paste("D:\\fin09\\P",1:n,"_0.C1",sep="")
os3<-paste("D:\\fin09\\Y",1:T*S,"_0.C1",sep="")
os4<-paste("D:\\fin09\\F",1:n,"_0.DAT",sep="")
#read in the point forecast into p0
p0<-matrix(rep(0,fy*J),ncol=J)
p0<-as.matrix(read.fwf(file=os1,widths=rep(10,J),skip=1))
#read in the target sample paths into ps
ps<-array(rep(0,fy*J*n),dim=c(fy,J,n))
for (k in 1:n) {
ps[,,k]<-as.matrix(read.fwf(file=os2[k],widths=rep(10,J),skip=1))
}
#read in the annual files for the computation of the conditional expectations
Y<-array(rep(0,N*J*T),dim=c(N,J,T))
for (k in 1:T)
Y[,,k]<-as.matrix(read.fwf(file=os3[k],widths=rep(10,J),skip=n+1))
```

}

A.2 fpath-output

```
#fpath-output.txt
#compute conditional expectations from PEP-output
#ASSUMES that fpath-input has been run and the results saved to an .RData file,
#such as f09.RData
#The following parameters were set:
#n = number of target paths and conditional forecast files
#N = number of files available for the computation of conditional expectations
#S = pick every Sth year
#T = number of forecast update years are required
#J = number of age-groups
#ne = number of vectors that form the neighborhood of the output path
#fy = number of forecast years in the source forecast made by PEP
#a = vector of N ones, used for the computation of distances
#e, st = end and beginning rows of blocks in output files
#os1-os4 = names of files for input and output
#p0 = point forecast
#ps = target sample paths
#Y = annual files
#START by clicking on the icon of the saved f09.RData
#the following are the output files as a 3-dimensional array
F<-array(rep(0,n*J*T*(T+1)/2),dim=c(T*(T+1)/2,J,n))
#the number of vectors that form the size of the of the neighborhood (may be different from ne)
ns1<-350
#the value of the scale parameter is
hs1<-7.4
#the parameters controlling the reduction in the effective size of the neighborhood
qs1<-0.95**c(=:T-2)
#k enumerates output sample paths
for (k in 1:n)
#this is the first block at jump-off
for (i in 1:T) {
F[i,,k]<-p0[i*S,]}
```

```
#select the neighborhood of size ns1
Di<-rowMeans(((aRa<-rand(Di)
Ne<-Y[Ra<=ns1,]
a1<-as.vector(rep(1,ns1))
#now do the remaining blocks b = 2:T
for (b in 2:T) {
#form the neighborhood of the chosen sample path
#based on chi-squared distance to the target data available at time (b-1)*S,
#and form weights for kernel estimation of the means of the increments
D<-rowMeans(((a%*%t(ps[(b-1)*S,c(1:19),k])-Ne[,c(1:19),b-1])**2)%*%diag(
1)*S,c(1:19),k]))
sc<-median(D)
w<-exp(-(hs1/qs1[b-1])*D/sc)/sum(exp(-(hs1/qs1[b-1])*D/sc))
F1<-ps[(b-1)*S,,k]
for (j in st[b]:e[b])
F[j,,k]<-round(abs(F1+colSums(w*(Ne[,,j-st[b]+b]-Ne[,,j-st[b]+b-1]))))
F1<-F[j,,k]
}
}
write(t(F[,,k]),file=os4[k],ncolumns=J)
}
```

Aikaisemmin ilmestynyt ETLA Raportit-sarjassa (ennen ETLA Keskusteluaiheita)

Previously published in the ETLA Reports series (formerly ETLA Discussion Papers)

No 17 *Martti Kulvik – Marja Tähtinen – Pekka Ylä-Anttila*, Business and Intellectual Capital Development in Financial Riptide. Case Studies of Finnish Biotechnology and Pharmaceutical Companies Dispersing into Global Value Chains. 15.10.2013. 82 p.

No 18 *Olavi Rantala*, Postitoiminnan kehitys vuoteen 2020. 18.11.2013. 22 s.

No 19 *Heli Koski*, Yleispalveluvelvoitteen merkitys postin kannattavuudelle. 18.11.2013. 15 s.

No 20 *Tuomo Virkola*, Exchange Rate Regime, Fiscal Foresight and the Effectiveness of Fiscal Policy in a Small Open Economy. 3.3.2014. 62 p.

No 21 *Ville Kaitila – Tuomo Virkola*, Openness, Specialisation and Vulnerability of the Nordic Countries. 27.3.2014. 25 p.

No 22 *Mika Maliranta – Niku Määttänen*, Innovation, Firm Risk and Industry Productivity. 1.4.2014. 14 p.

No 23 *Olavi Rantala*, Saksan ja muun euroalueen kilpailukyvyn ero eurokriisin taustalla. 1.4.2014. 23 s.

No 24 *Juha Alho*, Nonparametric Estimation of Conditional Expectations for Sustainability Analyses. 25.8.2014. 19 p.

No 25 *Cinzia Alcidi – Daniel Gros*, Implications of EU Governance Reforms: Rationale and Practical Application. 6.5.2014. 26 p.

No 26 *Antti Suvanto – Kimmo Virolainen*, Mihin pankkiunionia tarvitaan? 7.5.2014. 21 s.

No 27 *Topias Leino – Jyrki Ali-Yrkkö*, How Does Foreign Direct Investment Measure Real Investment by Foreign-owned Companies? Firm-level Analysis. 15.5.2014. 25 p.

No 28 *Timo Nikinmaa*, Kone- ja metallituoteteollisuuden visio 2025. 23.5.2014. 52 s.

No 29 *Antti Pelkonen – Duncan A. Thomas – Terttu Luukkonen*, Project-based Funding and Novelty in University Research – Findings from Finland and the UK. 12.6.2014. 18 p.

No 30 *Antti Kauhanen*, Tulevaisuuden työmarkkinat. 6.8.2014. 16 s.

No 31 *Joonas Tuhkuri*, Big Data: Google-haut ennustavat työttömyyttä Suomessa. 14.8.2014. 31 s.

Sarjan julkaisut ovat raportteja tutkimustuloksista ja väliraportteja tekeillä olevista tutkimuksista.

Julkaisut ovat ladattavissa pdf-muodossa osoitteessa: www.etla.fi » julkaisut » raportit

Papers in this series are reports on research results and on studies in progress.

Publications in pdf can be downloaded at www.etla.fi » publications » reports

ETLA
Elinkeinoelämän tutkimuslaitos
The Research Institute of the Finnish Economy
Lönnrotinkatu 4 B
00120 Helsinki

Puh. 09-609 900
Fax 09-601 753
www.etla.fi
etunimi.sukunimi@etla.fi

ISSN-L 2323-2447, ISSN 2323-2447, ISSN 2323-2455 (Pdf)