Modeling Convergence Dynamics: A Distance-Based Approach*

Constantina Kottaridi† Dimitrios D. Thomakos‡

Department of Economics

University of Peloponnese

Tripolis Campus, 22100 Greece

November 19, 2007

Extended Abstract

1 Introduction

In this paper we present an alternative approach for exploring convergence dynamics based on the use of distances. Our methodology is underlined by an exploratory data analysis (EDA) spirit in that the proposed empirical models are not derived from a theoretical economic model, but use the very notion of convergence as a reduction in some form of "distance" (e.g. output growth and its steady state) between a pair or group of variables of interest. Hence our approach potentially has applicability in many situations where convergence dynamics are to be examined, without the prior need of a theoretical model . Within the context of a distance-based convergence model we are able to provide estimates for the "speed of convergence", half-life and illustrate the fitted convergence path. We propose a variety of models for doing so, in-

*An earlier version of the paper was presented at the 5th CRETE conference, July 2006, at Rethymno, Greece. We would like to thank the conference participants for useful comments. All errors are ours. This is *very* preliminary material: please do not quote.

†Email: kottarid@uop.gr

‡Email: thomakos@uop.gr

cluding single equation, pool and panel equations and nonparametric equations for getting the above estimates. For most cases of practical empirical interest we are also able to provide an assessment of convergence based on the distribution of pairwise differences between variables of interest. We illustrate these ideas on the next section.

The literature on modeling convergence is really large and expanding rapidly. We do not provide a comprehensive review in this abstract but we do connect our approach with some related work. In the context of time series models convergence tests have been proposed based on the use of differences but *not distances*, see for example Quah (1992) Bernard (1992), and Bernard and Durlauf (1995). In such tests convergence is found if the difference between two variables of interest is a stationary time series. When combining time series and cross-sectional data in a panel a variety of papers use a number of popular panel unit root tests to test for convergence, see for example Evans and Karras (1996), Evans (1998) for representative applications in the context of growth convergence.[1] An interesting, recent technical reference on the use of such models, their pitfalls and potential remedies is given Phillips and Sul (2003). The use of cross-sectional variances (a measure of dispersion and thus "distance") was explicitly considered in Evans (1996). In a number of papers the issue of convergence is addressed with the use of predictive densities and distributional dynamics, differing in the specific methods used. See for example Azariadis and Stachurski (undated) on the use of stochastic kernels in assessing convergence using parametric estimates from a structural model as inputs in computing the kernels; Maasoumi, Racine and Stengos (undated) on the use of entropy distance between cross-country distributions for assessing convergence; Pittau and Zelli (undated) again on assessing the evolution of cross-country distributions via kernel densities; Canova (2004) using a predictive density approach to examine the existence of convergence clubs. Rappaport (2000) provides arguments in favor of and a framework for time-varying speed of convergence. Nahar and Inder (2002) examine convergence in a context that has some similarities to our work in the use of distances in assessing convergence but is considerably more narrow than the methodology we propose in this paper. Finally, a recent paper that uses pairwise distance measures, as we do, but in a different empirical context than ours is Pesaran (2007).

[1]Most related literature on convergence addresses growth and income convergence.

2 Distance-based Convergence Dynamics

Consider a group of p countries and, to begin with, a single variable of interest $X_i(t)$, for $i = 1, 2, \ldots, p$. In the case where there is a target variable for assessing convergence (e.g. a measure of a slowly-varying "steady state"), say $\bar{X}_i(t)$ then the pairwise distance measure between $X_i(t)$ and the target is defined as:

$$d_i^s(t) \stackrel{\text{def}}{=} |X_i(t) - \bar{X}_i(t)|^s \tag{1}$$

for $s \geq 1$ and the total distance measure for all p countries will be given by:

$$d_T^s(t) \stackrel{\text{def}}{=} \sum_{i=1}^{p} d_i^s(t) \Leftrightarrow d_T(t) \stackrel{\text{def}}{=} \left[\sum_{i=1}^{p} d_i^s(t)\right]^{1/s} \tag{2}$$

The above equation is the well-known Minkowski metric, where $s = 1$ corresponds to the Manhattan metric and $s = 2$ corresponds to the Euclidean metric. Other types of distance can also be used, for example the maximum distance measure $d_T^{\max}(t) \stackrel{\text{def}}{=} \max_i d_i^1(t)$.

Suppose next that there is no target variable available with respect to which we can assess distance, so that we can only work with the pairwise distances between members of the group. Let us define by $d_{ij}^s(t)$ the pairwise distance between member i and member j, i.e. $d_{ij}^s(t) \stackrel{\text{def}}{=} |X_i(t) - X_j(t)|^s$. We can arrange these pairwise distances in the $(p \times p)$ symmetric distance matrix $\boldsymbol{D}(t)$ given by:

$$\boldsymbol{D}(t) \stackrel{\text{def}}{=} \begin{bmatrix} d_{11}(t) & d_{12}(t) & \ldots & d_{1p}(t) \\ d_{21}(t) & d_{22}(t) & \ldots & d_{2p}(t) \\ \vdots & \vdots & \vdots & \vdots \\ d_{p1}(t) & d_{p2}(t) & \ldots & d_{pp}(t) \end{bmatrix} \tag{3}$$

and note that it has $q \stackrel{\text{def}}{=} p(p-1)/2$ distinct, non-zero elements, lying below or above the main diagonal. These unique elements will be useful in the analysis that follows so we formally defined them in the $(q \times 1)$ vector $\boldsymbol{d}(t)$ as:

$$\boldsymbol{d}(t) \stackrel{\text{def}}{=} \boldsymbol{S}^\top \mathsf{vech}(\boldsymbol{D}(t)) \tag{4}$$

where the $\mathsf{vech}(\boldsymbol{A})$ operator stacks the $q + p$ elements of the symmetric matrix $\boldsymbol{A}$ lying on and below the main diagonal and where the $((q + p) \times q)$ selection matrix $\boldsymbol{S}$ contains the columns of the $((q+p) \times (q+p))$ identity matrix that correspond to the non-zero elements in $\mathsf{vech}(\boldsymbol{D}(t))$. We can construct a total distance (scalar) measure using the elements of $\boldsymbol{d}(t)$ as before:

$$d_P^s(t) = \sum_{i \neq j}^{q} \left[d_{ij}^s(t)\right] \Leftrightarrow d_P(t) \stackrel{\text{def}}{=} \left[\sum_{i \neq j}^{q} d_{ij}^s(t)\right]^{1/s} \tag{5}$$

However, the vector of distinct distances $\boldsymbol{d}(t)$ can be used to formulate a multivariate and/or a panel model for analyzing convergence.

We note that the use of distances implies that we use a non-linear transformation in the differences between a pair of variables or a variable and its target. This is important as it allows us to consider potentially simpler models for assessing convergence. Specifically, if convergence takes place over a period of time then we must observe a decrease in the distance and the rate of change in the distance measure we are using should be negative, or:

$$\begin{array}{lcl} \lim_{t\to\infty}\langle d(t)\rangle & = & \gamma \to 0 \\ \partial\langle d(t)\rangle/\partial t & < & 0 \end{array} \tag{6}$$

where $d(t)$ is any of the distance measures we considered before and where $\langle X\rangle$ denotes the expected value of a random variable X. Note that we allow for a possibly non-zero limiting value γ since it is not necessary to have achieved zero distance to claim that convergence has taken place. Since convergence is associated with a decrease in distance the class of "decay" models can be appropriate for modeling convergence dynamics. The simplest such model can be derived from a standard differential equation for $\langle d(t)\rangle$ given by:

$$\frac{\partial\langle d(t)\rangle}{\partial t} \stackrel{\text{def}}{=} -\beta(t)\left[\langle d(t)\rangle - \gamma\right] \tag{7}$$

where $\beta(t)$ is the (possibly varying) "speed of convergence" (equivalent, the rate of decay in the distance) and γ is the limiting value as $t \to \infty$ as given in equation (6). Note that we have $\langle d(t)\rangle \geq \gamma$ for all t and that convergence takes place when $\beta(t) > 0$ for all t - the model however does not preclude the possibility of divergence, i.e. $\beta(t) \leq 0$. In this equation the "speed of convergence" is proportional to the difference of the distance at time t from the limiting value of γ and we can re-write the time derivative as:

$$\frac{\partial\langle d(t)\rangle}{\partial t} \cdot \frac{1}{\langle d(t)\rangle - \gamma} = -\beta(t) \tag{8}$$

If we assume a constant speed of convergence $\beta(t) = \beta$, as does most of the literature, then the solution of equation (7) takes the form of a two-parameter exponential decay equation:

$$\langle d(t)\rangle = \gamma + [d(0) - \gamma]\exp(-\beta t) \tag{9}$$

which is expressible as a function of time only. It is well known that a differential equation like equation (7) with constant coefficient β has an equivalent autoregressive solution given by:

$$\langle d(t)\rangle = \phi_0 + \phi_1\langle d(t-1)\rangle \tag{10}$$

where $\phi_0 \stackrel{\text{def}}{=} \gamma\,[1 - \exp(-\beta)]$ and $\phi_1 \stackrel{\text{def}}{=} \exp(-\beta)$. However, the solution in equation (9) is much more preferable from an empirical point of view: one does not have to dwell on the time series properties of the distance series $d(t)$ when using equation (9), as is simply a non-linear function of time.

Converting equation (8) into a stochastic, estimable equation, by adding a series of random deviations $u(t)$ around the deterministic dynamics, we can write it in a couple of alternative forms, to allow for single or pool/panel estimation. The properties of $u(t)$ are to be determined within the context of convergence. For single equation estimation we can use $d(t) = d_T(t)$ from equation (2) or $d(t) = d_P(t)$ from equation (5) and write:

$$d(t) = \gamma + [d(0) - \gamma]\exp(-\beta t) + u(t) \tag{11}$$

For short time series we may want to pool together the individual differences of equation (1) or the pairwise differences of equation (3) and consider the model as:

$$\begin{aligned} d_i^s(t) &= \gamma_i + [d_i(0) - \gamma_i]\exp(-\beta_i t) + u_i(t) \\ d_{ij}^s(t) &= \gamma_{ij} + [d_{ij}(0) - \gamma_{ij}]\exp(-\beta_{ij} t) + u_{ij}(t) \end{aligned} \tag{12}$$

Placing restrictions on the individual coefficients we can differentiate between panel and pool estimation. We have three different options per equation: for example, when using $d_i^s(t)$, we can vary both γ_i's and β_i's, vary only the γ_i's and have a common β, or have both γ and β common for all i's (and similarly for the γ_{ij}'s and β_{ij}'s).[2] Estimation of both equation

[2]Note that by using the solution of equation (9) and not the autoregressive representation of equation (10) we can avoid all shorts of problems associated with the estimation of dynamic panels.

(11) and equation (12) is straightforward by non-linear least squares or non-linear weighted least squares, depending on the specification. If the number of countries p is large then some restrictions would need to be placed in either the γ's or the β's to reduce the number of free coefficients and achieve or speed-up the numerical optimization. After estimation we can easily obtain the fitted "convergence path" and obtain immediate visual estimates of half-lives.

If the conditions of equation (6) are satisfied, and we thus have convergence, we would expect that the random deviations given by $u(t)$ would be transitory and their (unconditional) variance would probably decrease over time. It is not immediately apparent whether the $u(t)$ series should be stationary, if we allow for the possibility that the magnitude of the deviations from the convergence path is decreasing over time. Based on our empirical observations we propose the following as a potentially plausible structure for the $u(t)$ series:

1. $u(t)$ is mean-reverting around its unconditional mean of zero.
2. $u(t)$ can exhibit strong persistence at low lags that, however, decays exponentially fast.
3. $u(t)$ can exhibit low-frequency cyclical behavior - we make no claims as to whether there are "real" underlying cycles or not.
4. $u^2(t)$ or $|u(t)|$ can exhibit a negative trend, accounting for a decreasing variance over time.

The above structure is, as noted, completely based on empirical observations and the futures of our data although it appeared in practically all grouping combinations that we tried.

In addition to the parametric models discussed above we can make good use of the pairwise differences $d_{ij}^{s}(t)$ by estimating their empirical distributions at two distinct time periods. The number of distinct pairwise differences for p countries is $q = p(p-1)/2$. For sufficiently large p, e.g. $p > 15$ we obtain $q > 100$ observations of pairwise differences. These observations can be used in computing the empirical distributions at two different time periods (e.g. the beginning and end of the sample) and then compared, either visually or by formal statistical tests for their differences. Specifically, let $\{d_r(t_j)\}_{r=1}^{q} \stackrel{\text{def}}{=} \{d_{12}(t_j), d_{13}(t_j), \ldots, d_{1p}(t_j), \ldots, d_{p-1,p}(t_j)\}$ denote the $(q \times 1)$ sequence of pairwise distances at period t_j and compute the empirical distribution function $\widehat{F}(\delta, t_j)$ by:

$$\widehat{F}(\delta, t_j) \stackrel{\text{def}}{=} \frac{1}{q} \sum_{r=1}^{q} I(d_r(t_j) \leq \delta) \tag{13}$$

where $I(A)$ is the indicator function of the set A. If convergence takes place then, for $t_1 \gg t_0$ being two different time periods sufficiently apart in time, we should observe that:

$$\widehat{F}(\delta, t_1) \geq \widehat{F}(\delta, t_0) \quad \forall\, \delta \tag{14}$$

The condition in the above equation can be easily visualized and checked but we can also apply formal tests for distributional distances, akin to the approach of Maasoumi, Racine and Stengos (undated). Note that this condition is similar to conditions applied in empirical distributions in the context of stochastic dominance

3 Summary of Empirical Illustrations

Our first application of the suggested methodology is in the context of European inflation convergence. Using monthly data from 1991/1994 to 2005 we consider the annual inflation rates of the EU-25 countries (and a number of sub-groups). We apply our parametric models using the distances from a common target variable $\bar{X}(t)$, taken to be the average annual inflation of France, Germany and Netherlands for each month in the sample. We then consider the distributional differences using pairwise differences.

Our second application is in the context of income convergence within several OECD countries. Using annual data from 1970/1975 to 2005 on real GDP per capita we apply our parametric models using both pairwise distances and distances from a common target. We take the common target to be either the US real GDP or the average real GDP for each month in the sample. We also consider the distributional differences using the pairwise differences.

We next present a set of representative results from our first application. We consider the group of all EU-25 countries. Estimation results based on the parametric model of equation (11), using the Manhattan distance metric, are given below. We provide the estimates of the (β, γ) parameters and a number of statistics for the residual series $\widehat{u}(t)$ to illustrate their properties. The estimated equations are given by:

$$\begin{aligned} d_T(t) &= 23.85 + [d(0) - 23.85] \exp(-0.03t) + \widehat{u}(t) \\ \widehat{u}(t) &= 1.16\widehat{u}(t-1) - 0.24\widehat{u}(t-2) + \widehat{e}(t) \\ |\widehat{u}(t)| &= 10.13 - 0.05t + \widehat{\eta}(t) \end{aligned} \tag{15}$$

where all estimates are significant at the 5% level. The "speed of convergence" is estimated at 3% while the limiting value is estimated at 23.85. Note that this is *not* a large value: since we are using the total distance series of equation (2) it corresponds to a sum over the 25 countries. We should divide it by 25 to get it appropriately scaled, i.e. to obtain 0.954. For comparison, note that the estimates of "speed of convergence" and the limiting value in the context of pooled estimation using equation (12) are 3% and 1.18 respectively, so the estimates are practically the same. In Figure 1 we plot the actual distances and the fitted values corresponding to the convergence path. We can easily see from the plot that the half-life is about 3 years.

The analysis of the residual series $\widehat{u}(t)$ shows that they follow the properties we outlined in the previous section. An $AR(2)$ model with real distinct roots adequately captures the dynamic properties of the series. The series, plotted in Figure 2, does exhibit a cyclical pattern which cannot be accounted by the fitted AR model as its roots are not complex. Figure 3 has the spectral density of the series which does exhibit a marked peak at frequency 0.03. A regression of the absolute values of the residuals (as a proxy for their variance) against a time trend also shows that the variance of the residuals is decreasing over time. The actual and fitted values from this third regression are given in Figure 4. Finally, in Figure 5 we present the results from the estimation of the empirical distribution functions based on pairwise differences. It is evident that condition of equation (14) is clearly satisfied.

It is interesting to note that the pattern of results presented before appears both for various sub-groups in the context of the EU and in the analysis of the OECD data. We are currently working on producing a complete list with the related results.

References

[1] Azariadis, C. and J. Stachurski, undated, "A Forward Projection of the Cross-Country Income Distribution", mimeo.

[2] Bernard, A.B., 1992, "Empirical implications of the convergence hypothesis", mimeo.

[3] Bernard, A.B. and S.N. Durlauf, 1995, "Convergence in international output", *Journal of Applied Econometrics*, 10, 97-108.

[4] Canova, F. 2004, "Testing for Convergence Clubs in Income per Capita: A Predictive Density Approach", *International Economic Review*, 1, 49-77.

[5] Evans, P., 1996, "Using Cross-Country Variances to Evaluate Growth Theories", *Journal of Economic Dynamics and Control*, 20, 1027-1049.

[6] Evans, P. and G. Karras, 1996, "Convergence revisited", *Journal of Monetary Economics*, 37, 249-265.

[7] Maasoumi, E., Racine, J. and T. Stengos, undated, "Growth and Convergence: A Profile of Distribution Dynamics and Mobility", mimeo.

[8] Nahar, S. and B. Inder, 2002, "Testing Convergence in Economic Growth in OECD Countries", *Applied Economics*, 34, 2011-2022.

[9] Pesaran, M.H., 2007. "A Pair-wise Approach to Testing for Output and Growth Convergence", *Journal of Econometrics*, 138, pp. 312-355.

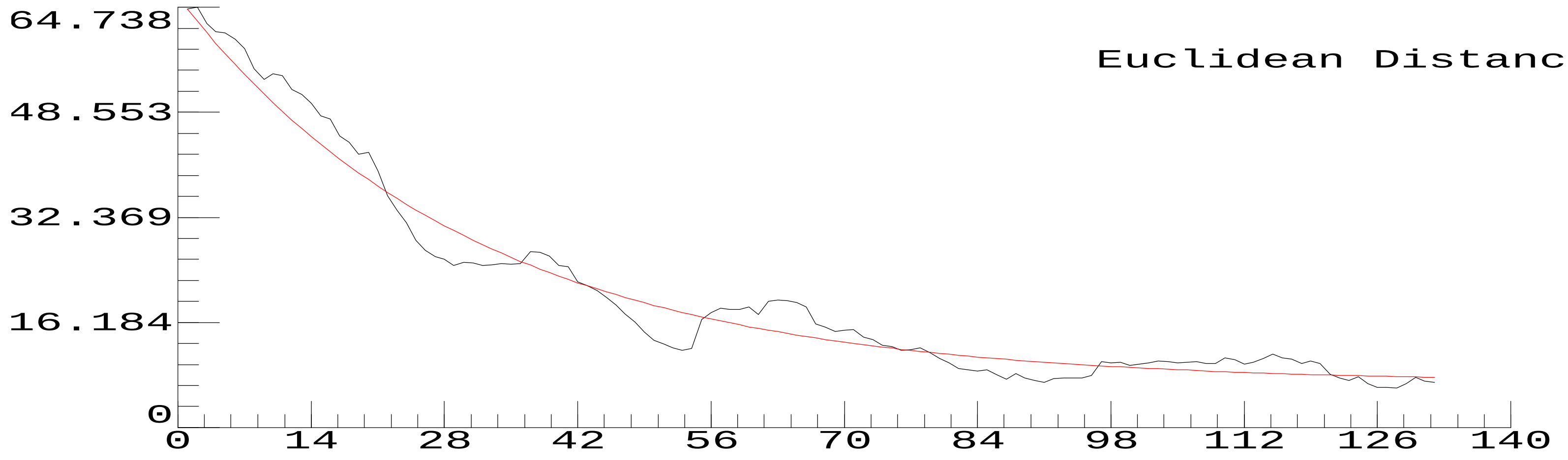
[10] Phillips, P.C.B. and D. Sul, 2003, "The Elusive Empirical Shadow of Growth Convergence", Cowles Foundation Discussion Paper no. 1398.

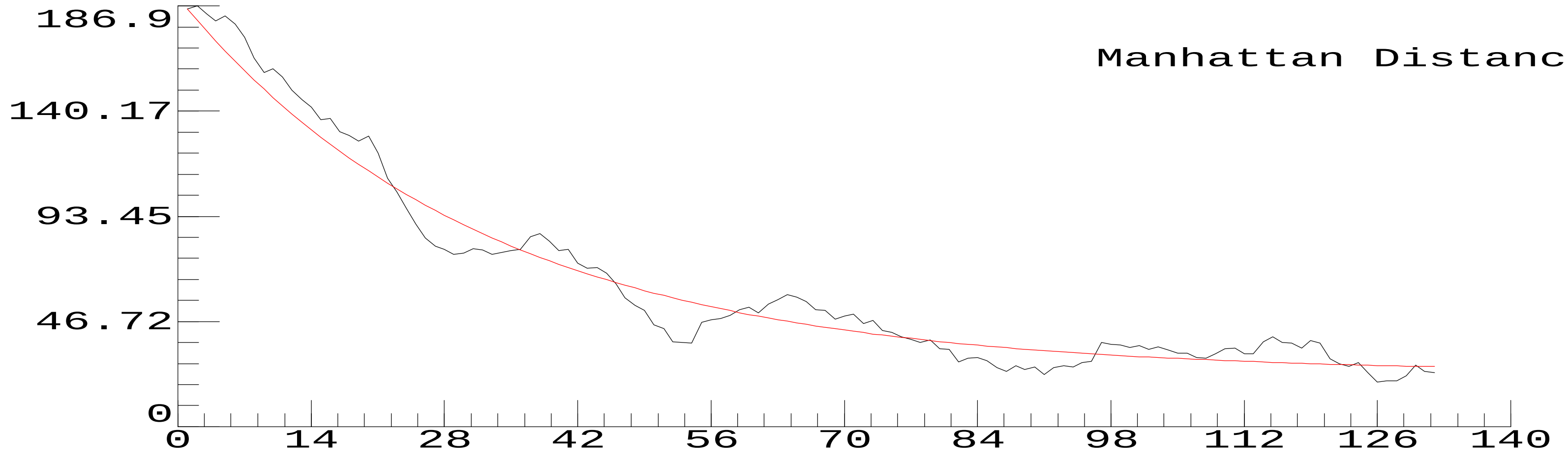
[11] Pittau, M.G. and R. Zelli, undated, "Empirical Evidence of Income Dynamics Across EU Regions", mimeo.

[12] Quah, D., 1992, "International patterns of growth: I. Persistence in cross-country disparities", mimeo.

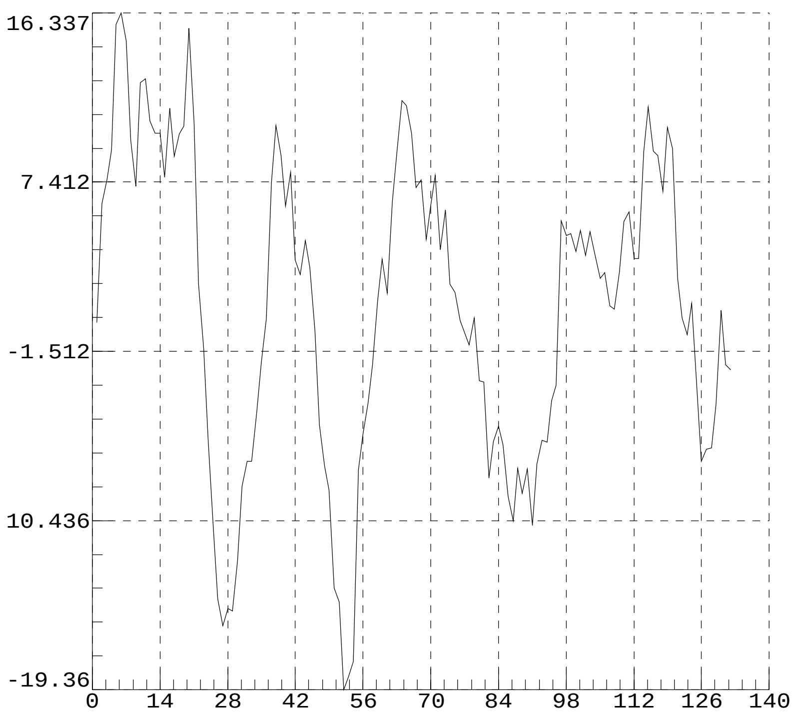
[13] Rappaport, J., 2000, "Is the Speed of Convergence Constant?", mimeo.

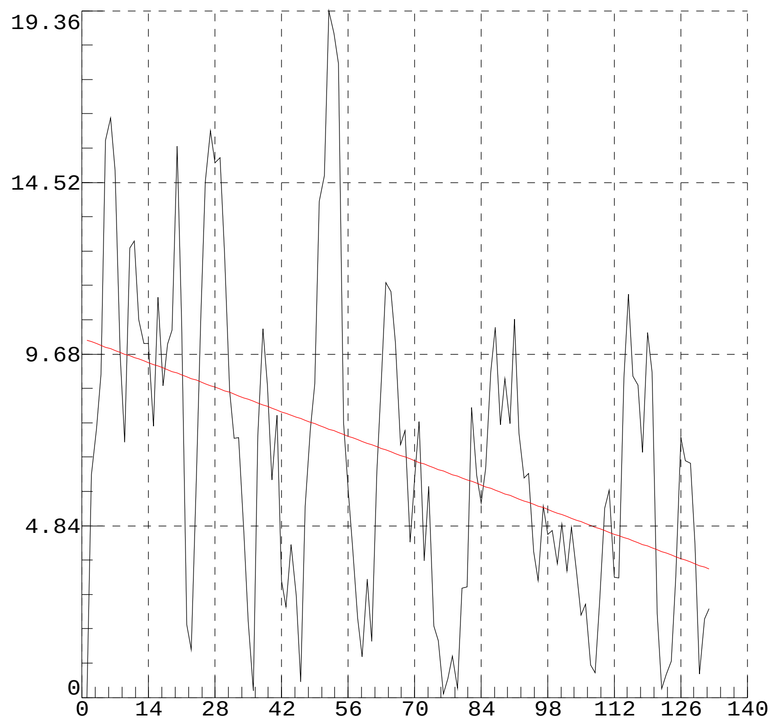
Distances and Convergence Paths: European Inflation, EU-25



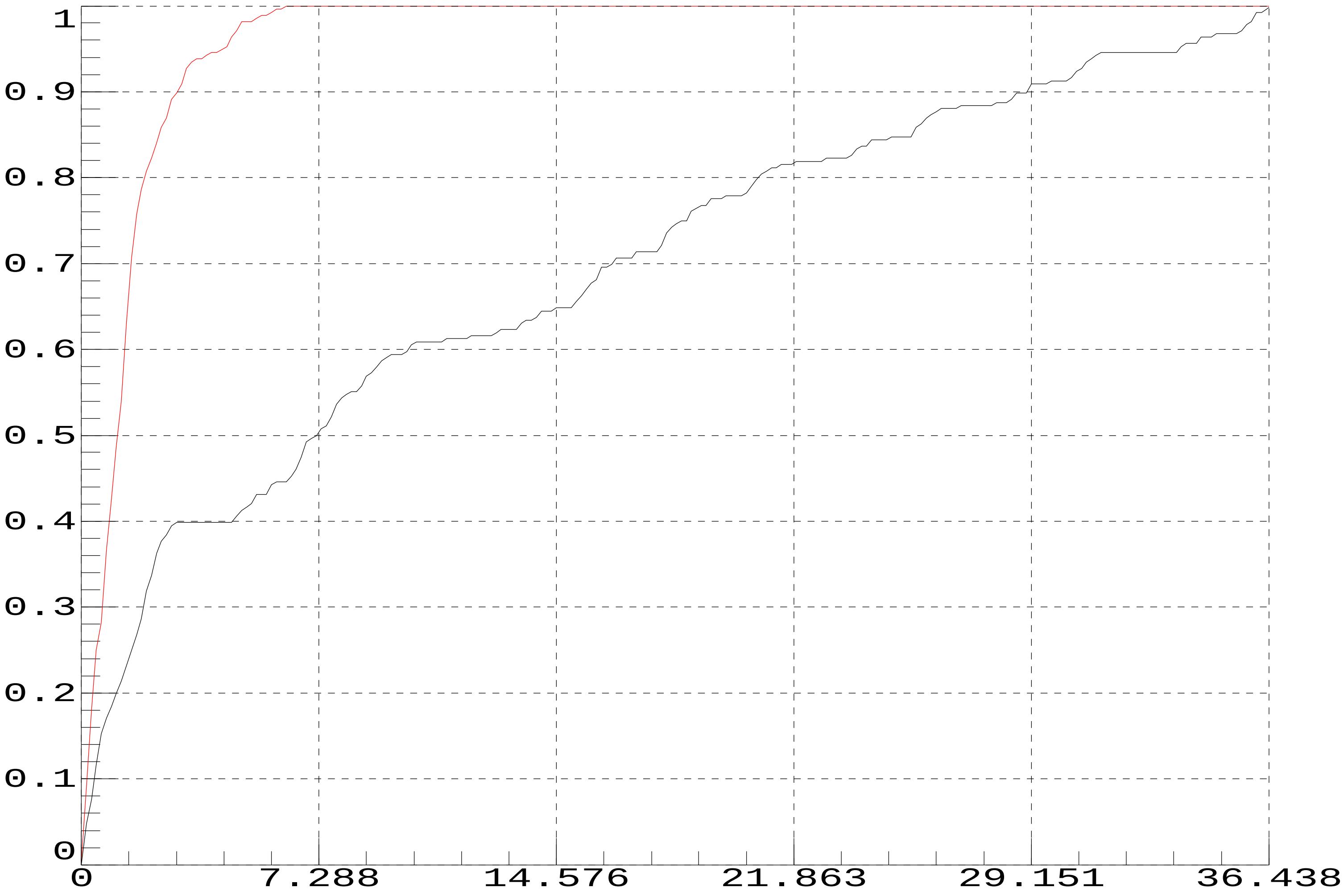


Residuals from Fitted Path and Absolute Residuals with Fitted Trend

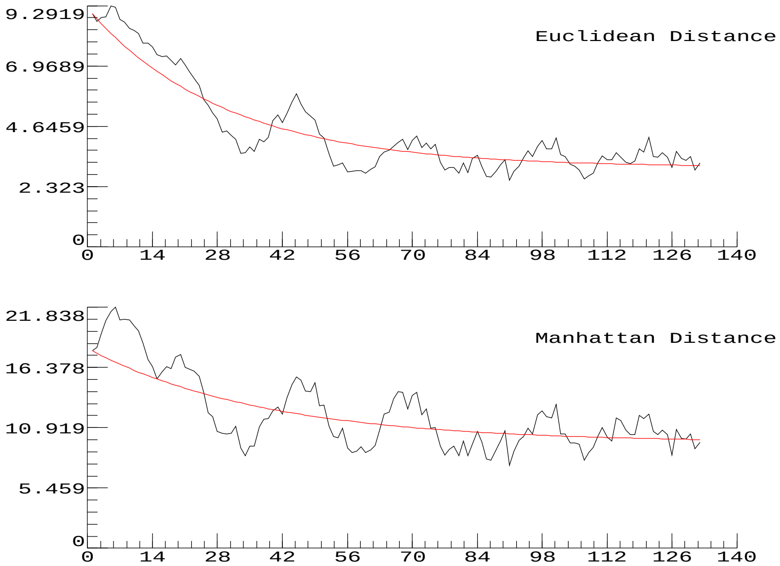


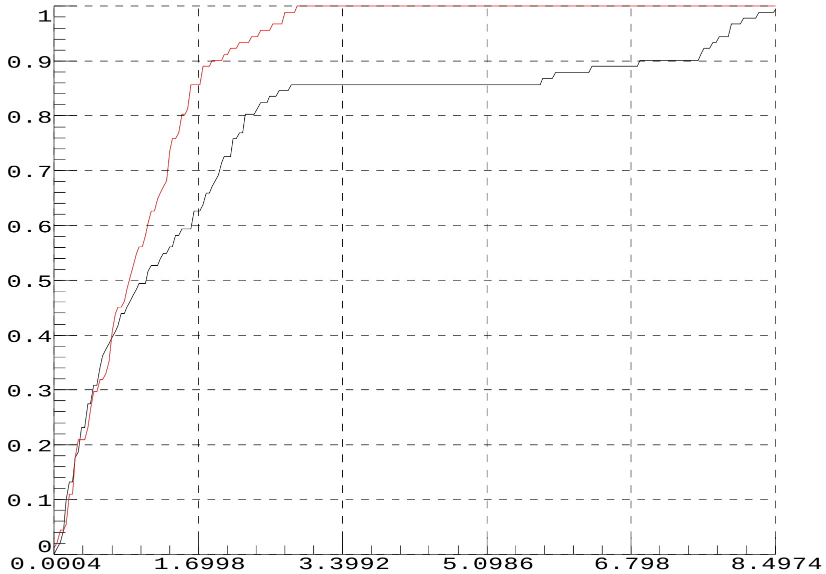


Empirical Distribution of Pairwise Differences: European Inflation, EU-25
Red line is end-of-sample, white line is beginning-of-sample

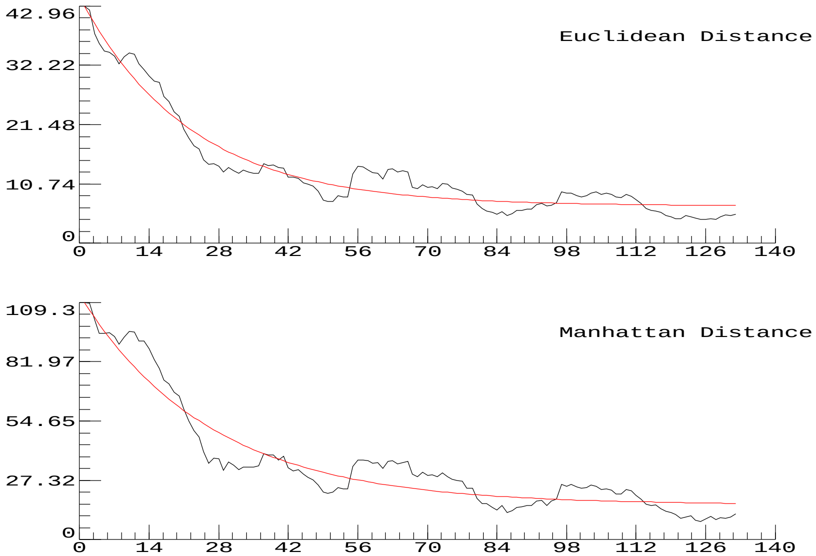


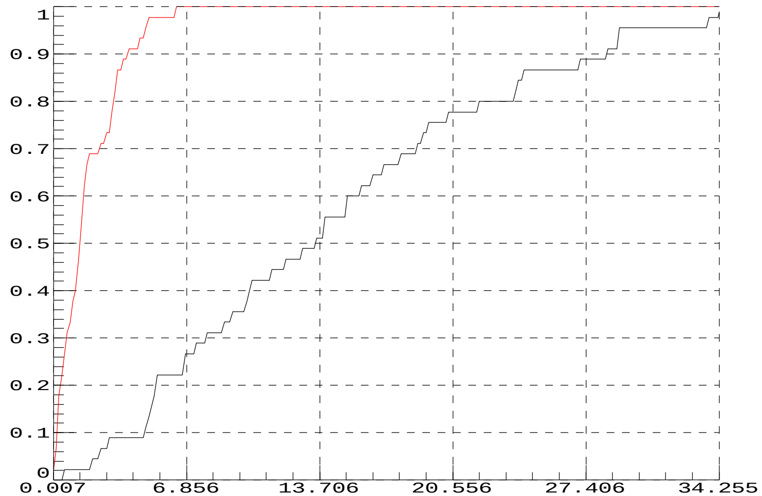
European Inflation, EU-15



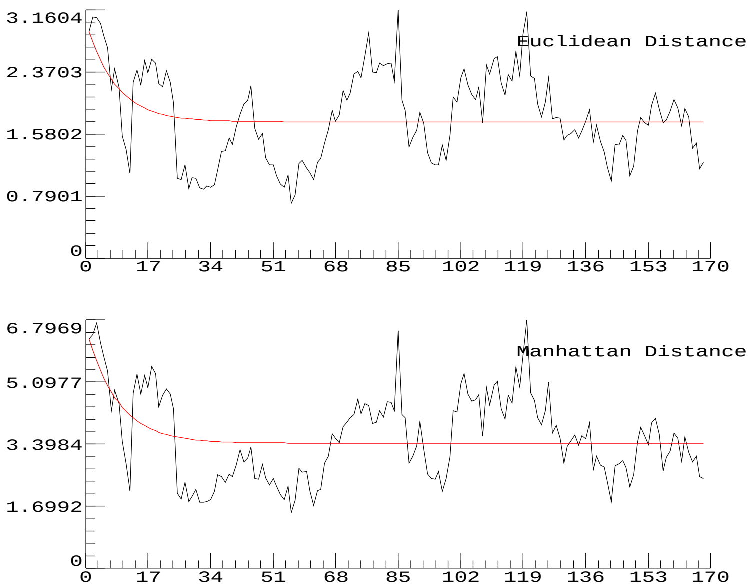


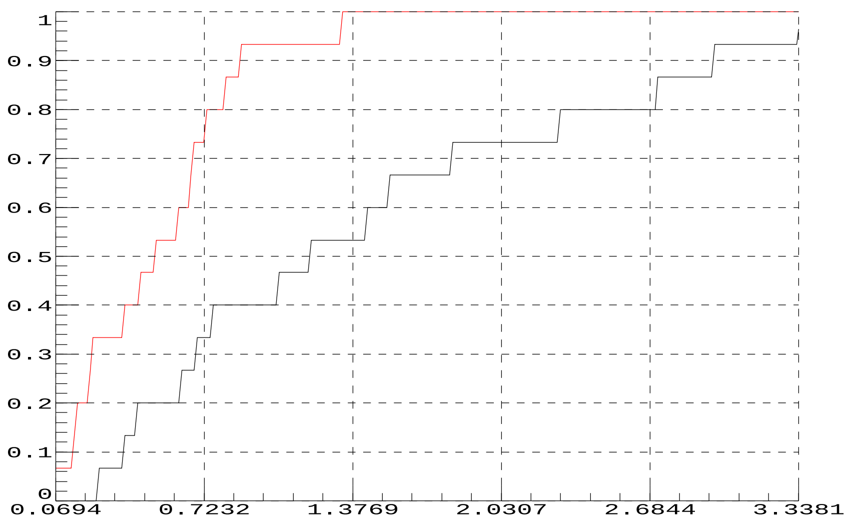
European Inflation, EU-10





European Inflation, Core





European Inflation, Med