Brabec, Marek und Komlos, John:

Spatial Convergence in Height in East-Central Europe, 1890-1910

Munich Discussion Paper No. 2007-3

Department of Economics
University of Munich

Volkswirtschaftliche Fakultät
Ludwig-Maximilians-Universität München

Online at https://doi.org/10.5282/ubm/epub.1358

Spatial Convergence in Height in East-Central Europe, 1890-1910

Marek Brabec, *Department of Biostatistics and Informatics, National Institute of Public Health., Srobarova 48, Praha 10, 10 042, Czech Republic,* mbrabec@szu.cz

and John Komlos, Department of Economics, University of Munich

Direct correspondence to John Komlos, Department of Economics, University of Munich, Ludwigstr. 33 / IV, 80539 Munich, Germany. Tel: +4989-2180-3169 John.Komlos@gmx.de

Abstract: We examine spatial convergence in biological well-being in the Habsburg Monarchy circa 1890-1910 on the basis of evidence on the physical stature of 21-year-old recruits disaggregated into 15 districts. We find that the shorter was the population in 1890 the faster its height grew thereafter. Hence, there was convergence in physical stature between the peripheral areas of the monarchy (located in today's Poland/Ukraine, Romania, and Slovakia) and the core (located in today's Austria, Czech Republic, and Hungary). The difference between the trend in the height of the Polish district of Przemysl and the Viennese trend was about 0.9 cm per decade in favor of the former. But the convergence among the core districts themselves was minimal or non-existent, whereas the convergence among the peripheral districts was more pronounced. Hence, spatial convergence took place exclusively within the peripheral areas, and between the peripheral regions and the more developed ones. The pattern is somewhat reminiscent of modern findings on convergence clubs in the global economy. However, the East-Central European pattern was the reverse of this modern finding: heights converged to the levels of the developed regions, but did not converge among the more developed regions themselves.

JEL: D63, I10, J10, N13, N930, 049

Keywords: spatial convergence, inequality, physical stature, economic growth, biological standard of living, Habsburg Monarchy, anthropometric history,

We explore spatial variability and convergence in the height of recruits in the Habsburg Monarchy for the recruitment years 1890-1910. All men were measured annually and uniformly in this period at age 21 and the distributions of physical stature were reported in eight bins for 15 military districts. The bins are open-ended at the ends of the distribution and the bins are not equidistant.

Estimates of the trends in Height

We first estimate heights using a censored normal model, saturated in time and district. In other words, we stratify on time • district combinations and fit separate (interval censored) Gaussian models for separate combinations, without any restriction on parameters (both means and variances). We assume independence among different strata (time • district combinations). The data we have are in the form of aggregate tables giving frequency of several height intervals for each individual time-district combination. That is, we have $C(I_{itj})$'s, as the counts of men being examined at 21 years at i-th district ($i = 1,\ldots,15$), year t ($t = 1890,\ldots,1910$, that is the recruitment or measurement year) whose height fell into the j-th interval I_{itj} ($j = 1,\ldots,8$). For a typical it-th time • district combination, we have (not equidistant) intervals $I_{it1} = (0,153)$, $I_{it2} = [153,155)$, $I_{it3} = [155,161)$, $I_{it4} = [161,166)$, $I_{it5} = [166,171)$, $I_{it6} = [171,176)$, $I_{it7} = [176,181)$, $I_{it8} = [181,\infty)$, all in centimeters.[1] The estimation is complicated by the fact that individual heights are not observed directly. Height observations can be regarded as *censored* (Cox and Oakes 1984). In fact, we have a mix of left-/right- and interval censoring which renders typical approximate methods unreliable (especially for variance estimation). On the other hand, maximum likelihood estimation (MLE) based on censoring is straightforward (Meeker and Duke (1981)). We use $C(I_{it1}),\ldots,C(I_{it8})$ data to get maximum likelihood based estimate of μ_{it},σ^2_{it} for each district-time combination or stratum. Effectively, for the underlying height distribution that we cannot observe directly (due to censoring) we assume the (saturated) model:

$$Y_{itk} = \mu_{it} + \varepsilon_{itk} \tag{1}$$

where Y_{itk} is the height of a man k, randomly chosen among those being recruited in district i and year t. Based on common practice of using Gaussian approximating for population height distribution[2], we assume that $\varepsilon_{itk} \sim N(0, \sigma_{it}^2)$, independently across i and t. Notice, that individual heights (Y_{itk} for individual k's) are not available. We have only counts for different height intervals (denoted by $C(I_{it1}), \ldots, C(I_{it8})$). These are corresponding to censored form of normal data generated by model (1). For count $C(I_{itl})$ in an interval I_{itl} with left point a_l and right point b_l, we have interval censoring in general. When $l = 1$ and $a_1 = -\infty$, we have left censoring and when $l = 8$ and $b_8 = \infty$, we have right censoring. Each of the intervals adds one term (of the general form: $C(I_{itl}).\log\left[\Phi\left(\frac{b_l - \mu_{it}}{\sigma_{it}}\right) - \Phi\left(\frac{a_l - \mu_{it}}{\sigma_{it}}\right)\right]$) in the likelihood function.

Numerical maximization of the loglikelihood produces not only (asymptotically efficient) MLEs of μ_{it} and σ_{it}^2, but also their asymptotic standard errors (via inversion of the information matrix).

Note that, although this is an estimable model, it is *very large* and not very parsimonious nor efficient. Its main advantage lies in the fact that it has minimal requirements and hence its results are not driven by preconceptions introduced by a priori theories. Hence it is useful for graphical expositions and as a baseline against which more restrictive models can be compared. Figure 1 shows the MLEs $\hat{\mu}_{it}$ as separate time trajectories for different districts.[3]

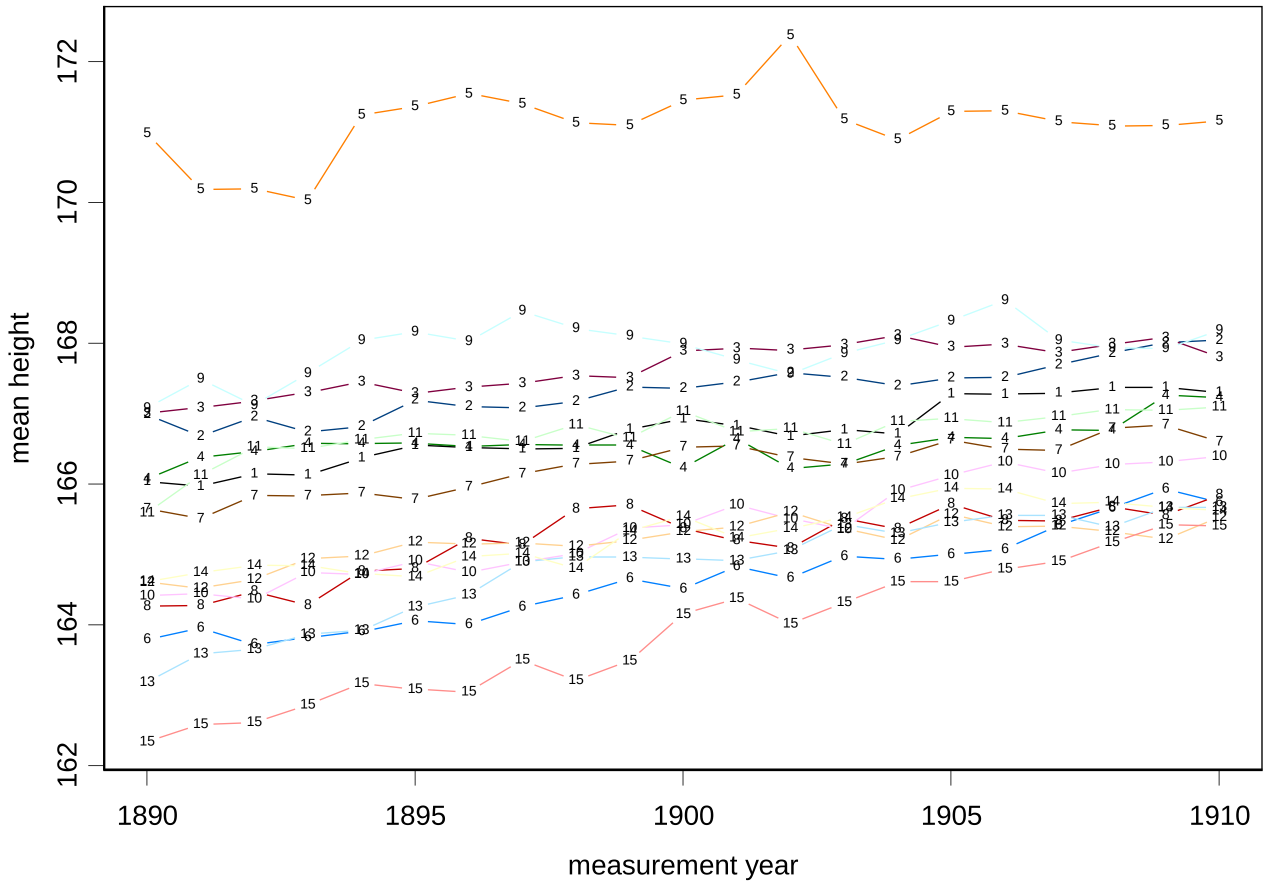

Figure 1 Estimated Height (cm) by 15 Districts of the Habsburg Monarchy. 1=Vienna, 2=Prague, 3=Gratz, 4=Innsbruck, 5= Zara (Croatia), 6=Lemberg (Galicia, Poland), 7=Budapest, 8=Hermannstadt (Transylvania, Romania), 9=Zagreb (Croatia), 10=Krakau (Galicia, Poland), 11=Josefstadt (Czech), 12=Pressburg (Slovakia), 13=Kaschau (Slovakia), 14=Temesvar (Transylvania, Romania), 15=Przemysl (Galicia, Poland).

To a first approximation, the trends appear roughly linear and not exponential as assumed in the Barro/Sala-i-Martin (1995) model. Hence, we use linear regression of mean height (from saturated model estimates) on time (measurement year-1890), weighted by reciprocal square of SEM[4] (obtained from saturated model (1)). This amounts to a (global) linear approximation to the potentially (slightly) nonlinear district-specific mean height trajectories and hence to a more restrictive model:

$$Y_{itk} = \beta_{0i} + \beta_{1i}(t - 1890) + \varepsilon_{itk} \qquad (2)$$

where Y_{itk} is the height of a man k, randomly selected in district i and year t and $\varepsilon_{itk} \sim N(0, \sigma_i^2)$ (independently across i and t) is assumed. Note that regression coefficients β_{0i}, β_{1i} as well as within-district variances σ_i^2 are allowed to be different for different districts, but are restricted to be constant over time. Due to this restriction (2) is much smaller and simpler (or more parsimonious and easier to interpret) than (1) in the sense of having 45, instead of 630 parameters to estimate. Histogram of the estimated intercepts from the linear regression Eq. (2) is given in

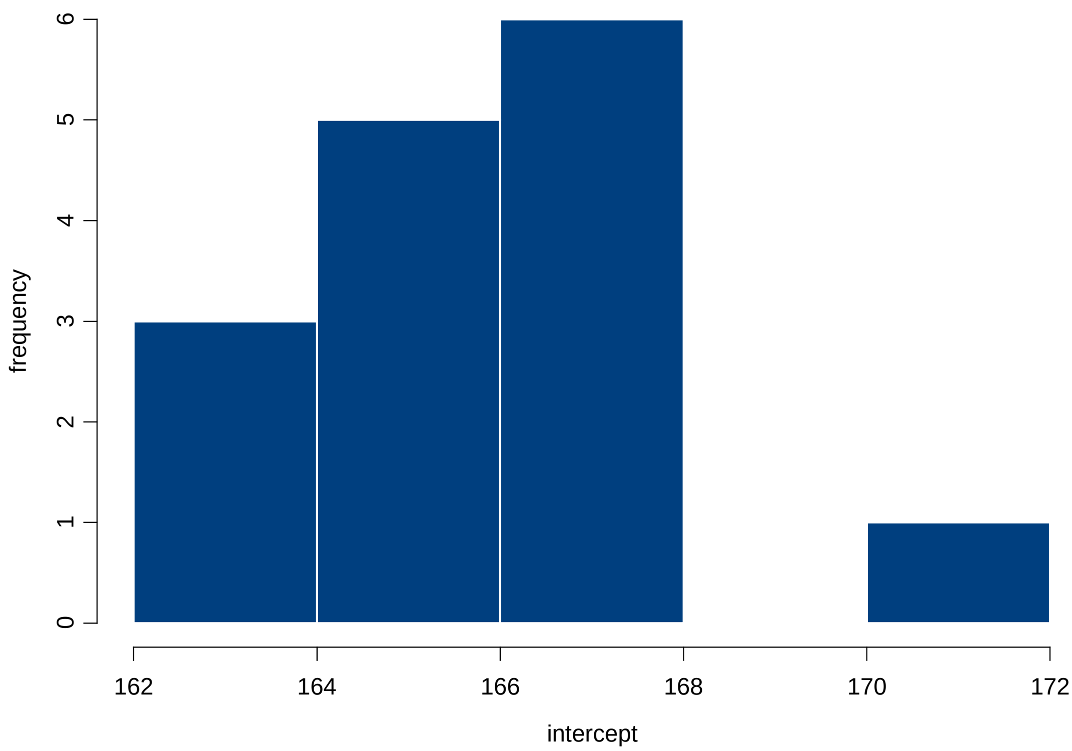

Figure 2. Distribution of district-specific intercepts (in cm) for 15 Districts of the Habsburg Monarchy

Table 1 shows estimates of model (2) parameters, together with their standard errors and p-values. All slopes are positive (as expected in connection with widely accepted positive secular trend) and all of them (except Zara's) are highly significant. At the estimated rates it would have taken the Polish district of Przemysl 43 years to catch up with the Austrian district of Vienna.

city	b0	b1	sb0	sb1	pb0	pb1	sigma
Vienna	166,033	0,07	0,063	0,005	0,000	0,000	0,141
Prague	166,742	0,059	0,056	0,005	0,000	0,000	0,128
Gratz	167,143	0,051	0,068	0,006	0,000	0,000	0,157
Innsbruck	166,275	0,03	0,098	0,008	0,000	0,002	0,224
Zara	170,896	0,026	0,215	0,018	0,000	0,150	0,475
Lemberg	163,561	0,107	0,072	0,006	0,000	0,000	0,162
Budapest	165,681	0,057	0,066	0,005	0,000	0,000	0,148
Hermannstadt	164,527	0,066	0,122	0,01	0,000	0,000	0,266
Zagreb	167,655	0,029	0,15	0,013	0,000	0,032	0,337
Krakau	164,281	0,11	0,067	0,006	0,000	0,000	0,157
Josefstadt	166,281	0,043	0,091	0,008	0,000	0,000	0,214
Pressburg	164,796	0,039	0,076	0,006	0,000	0,000	0,177
Kaschau	163,639	0,116	0,105	0,009	0,000	0,000	0,245
Temesvar	164,616	0,066	0,089	0,007	0,000	0,000	0,198
Przemysl	162,334	0,155	0,072	0,006	0,000	0,000	0,173

Table 1 Coefficient estimates for model (2) - 15 districts of the Habsburg Monarchy. (b0 is intercept, b1 is slope, sb0 and sb1 are their standard errors, pb0 and pb1 their p-values sigma is the residual standard deviation)

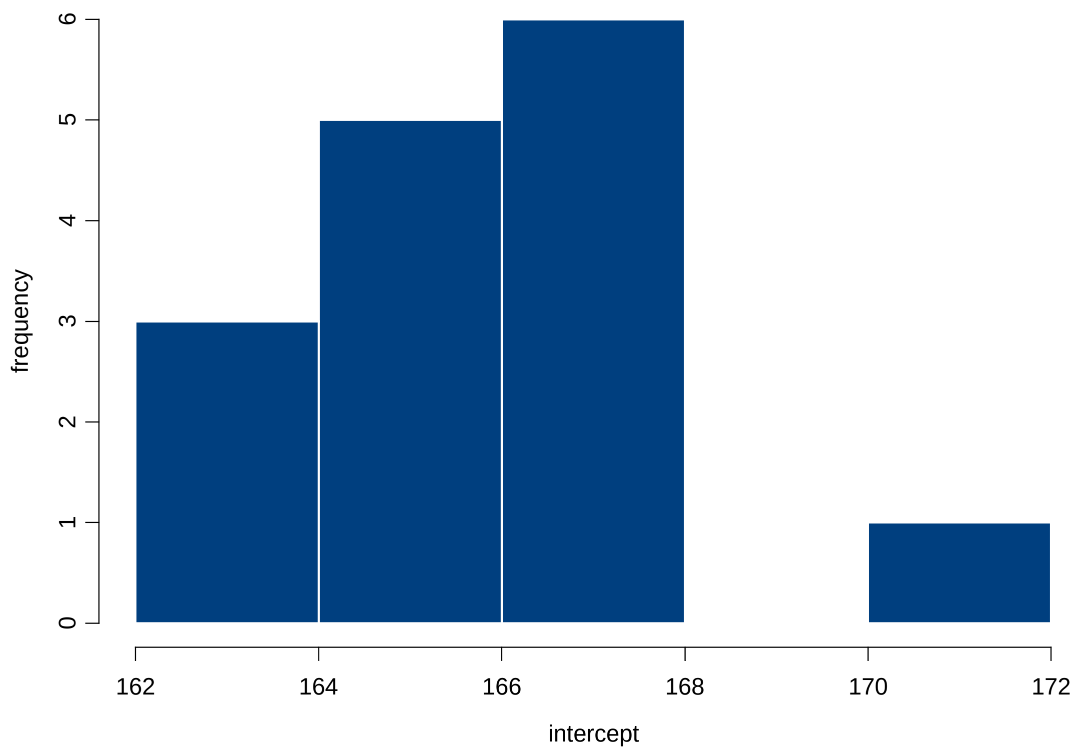

Figure 2. Distribution of district-specific intercepts (in cm) for 15 Districts of the Habsburg Monarchy

Table 1[Marek Bra1] shows estimates of model (2) parameters, together with their standard errors and p-values. All slopes are positive (as expected in connection with widely accepted positive secular trend) and all of them (except Zara's) are highly significant. At the estimated rates it would have taken the Polish district of Przemysl 43 years to catch up with the Austrian district of Vienna.

city	b0	b1	sb0	sb1	pb0	pb1	sigma
Vienna	166,033	0,07	0,063	0,005	0,000	0,000	0,141
Prague	166,742	0,059	0,056	0,005	0,000	0,000	0,128
Gratz	167,143	0,051	0,068	0,006	0,000	0,000	0,157
Innsbruck	166,275	0,03	0,098	0,008	0,000	0,002	0,224
Zara	170,896	0,026	0,215	0,018	0,000	0,150	0,475
Lemberg	163,561	0,107	0,072	0,006	0,000	0,000	0,162
Budapest	165,681	0,057	0,066	0,005	0,000	0,000	0,148
Hermannstadt	164,527	0,066	0,122	0,01	0,000	0,000	0,266
Zagreb	167,655	0,029	0,15	0,013	0,000	0,032	0,337
Krakau	164,281	0,11	0,067	0,006	0,000	0,000	0,157
Josefstadt	166,281	0,043	0,091	0,008	0,000	0,000	0,214
Pressburg	164,796	0,039	0,076	0,006	0,000	0,000	0,177
Kaschau	163,639	0,116	0,105	0,009	0,000	0,000	0,245
Temesvar	164,616	0,066	0,089	0,007	0,000	0,000	0,198
Przemysl	162,334	0,155	0,072	0,006	0,000	0,000	0,173

Table 1 Coefficient estimates for model (2) - 15 districts of the Habsburg Monarchy. (b0 is intercept, b1 is slope, sb0 and sb1 are their standard errors, pb0 and pb1 their p-values sigma is the residual standard deviation)

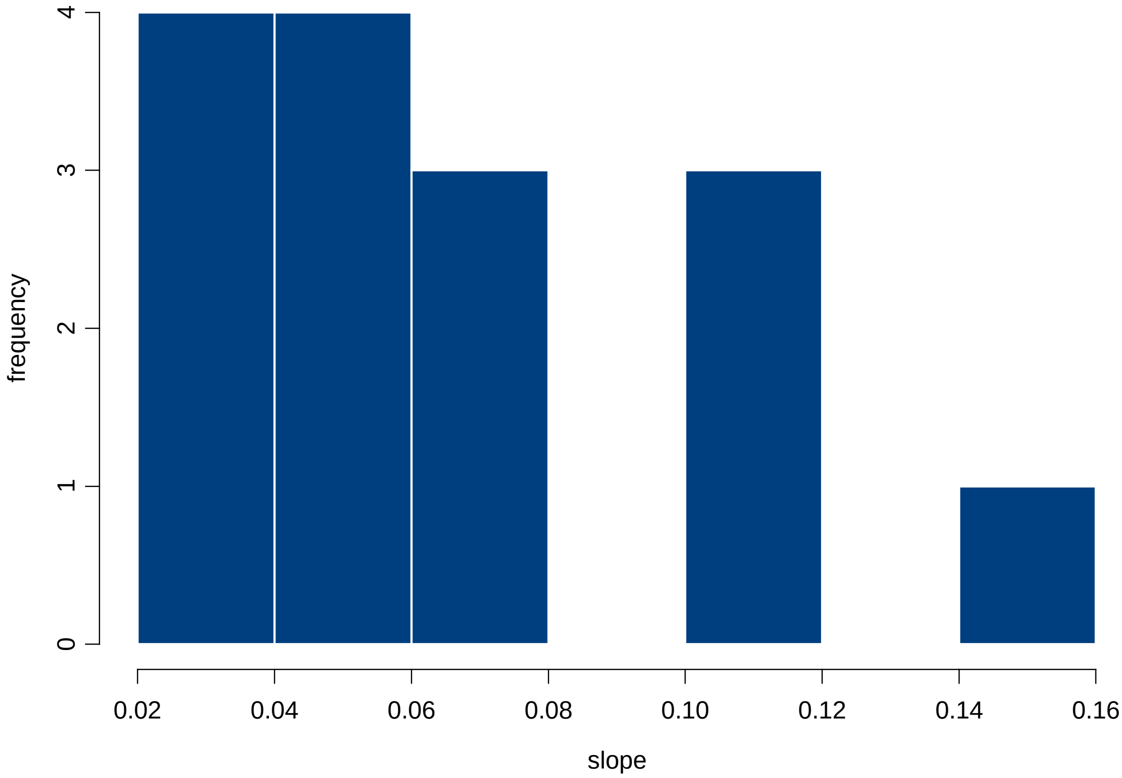

Figure 3 Distribution of district-specific slopes (in cm/year) for 15 Districts of the Habsburg Monarchy

Figure shows histogram of estimated slopes for height linear trend over time (from Eq. 2) (cm per annum) for 15 Districts of the Habsburg Monarchy in 1890. While most of the districts feature moderate growth (none of them shows negative growth or shrinkage), the range of growth rates is quite wide. Due to the parametrization used in (2), the intercepts correspond to the mean height at 1890 (at the start of the observation period and hence

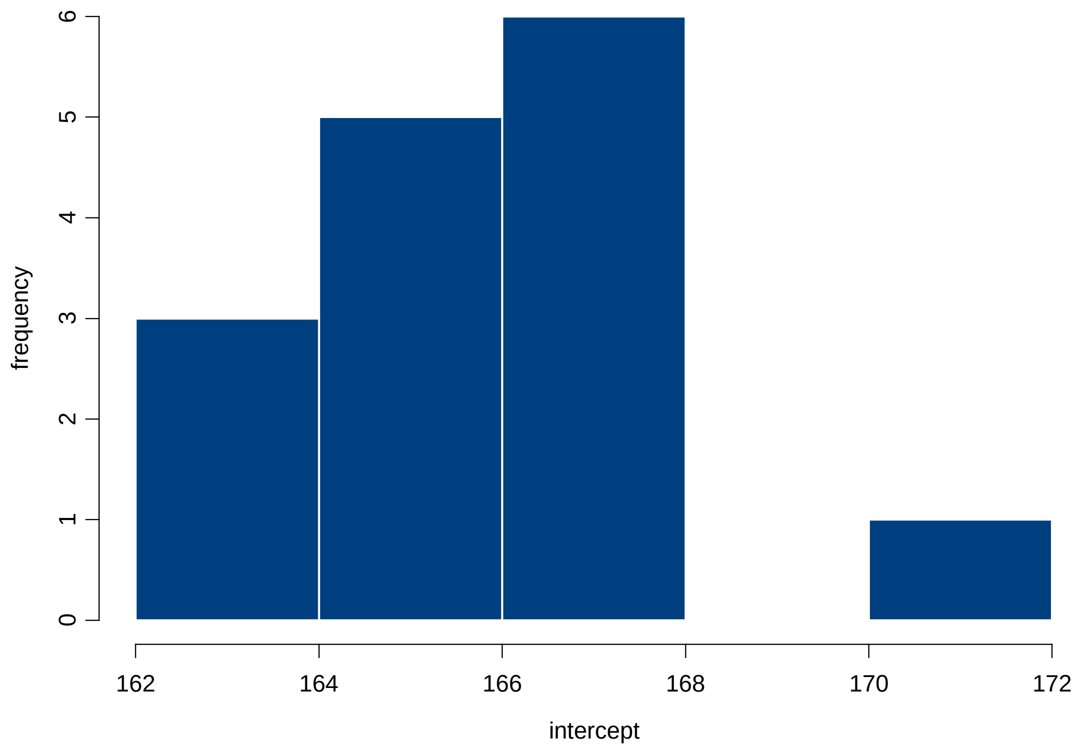

Figure 2. Distribution of district-specific intercepts (in cm) for 15 Districts of the Habsburg Monarchy

Table 1 shows estimates of model (2) parameters, together with their standard errors and p-values. All slopes are positive (as expected in connection with widely accepted positive secular trend) and all of them (except Zara's) are highly significant. At the estimated rates it would have taken the Polish district of Przemysl 43 years to catch up with the Austrian district of Vienna.

city	b0	b1	sb0	sb1	pb0	pb1	sigma
Vienna	166,033	0,07	0,063	0,005	0,000	0,000	0,141
Prague	166,742	0,059	0,056	0,005	0,000	0,000	0,128
Gratz	167,143	0,051	0,068	0,006	0,000	0,000	0,157
Innsbruck	166,275	0,03	0,098	0,008	0,000	0,002	0,224
Zara	170,896	0,026	0,215	0,018	0,000	0,150	0,475
Lemberg	163,561	0,107	0,072	0,006	0,000	0,000	0,162
Budapest	165,681	0,057	0,066	0,005	0,000	0,000	0,148
Hermannstadt	164,527	0,066	0,122	0,01	0,000	0,000	0,266
Zagreb	167,655	0,029	0,15	0,013	0,000	0,032	0,337
Krakau	164,281	0,11	0,067	0,006	0,000	0,000	0,157
Josefstadt	166,281	0,043	0,091	0,008	0,000	0,000	0,214
Pressburg	164,796	0,039	0,076	0,006	0,000	0,000	0,177
Kaschau	163,639	0,116	0,105	0,009	0,000	0,000	0,245
Temesvar	164,616	0,066	0,089	0,007	0,000	0,000	0,198
Przemysl	162,334	0,155	0,072	0,006	0,000	0,000	0,173

Table 1 Coefficient estimates for model (2) - 15 districts of the Habsburg Monarchy. (b0 is intercept, b1 is slope, sb0 and sb1 are their standard errors, pb0 and pb1 their p-values sigma is the residual standard deviation)

shows the distribution of "initial conditions" (across districts, i.e. spatially). Figure shows the distribution of secular trends (at 21 years) averaged across the observation period (1890-1910). Figure shows the histogram of the $\hat{\sigma}_i$'s as MLEs for the within-district (or residual) standard deviations.

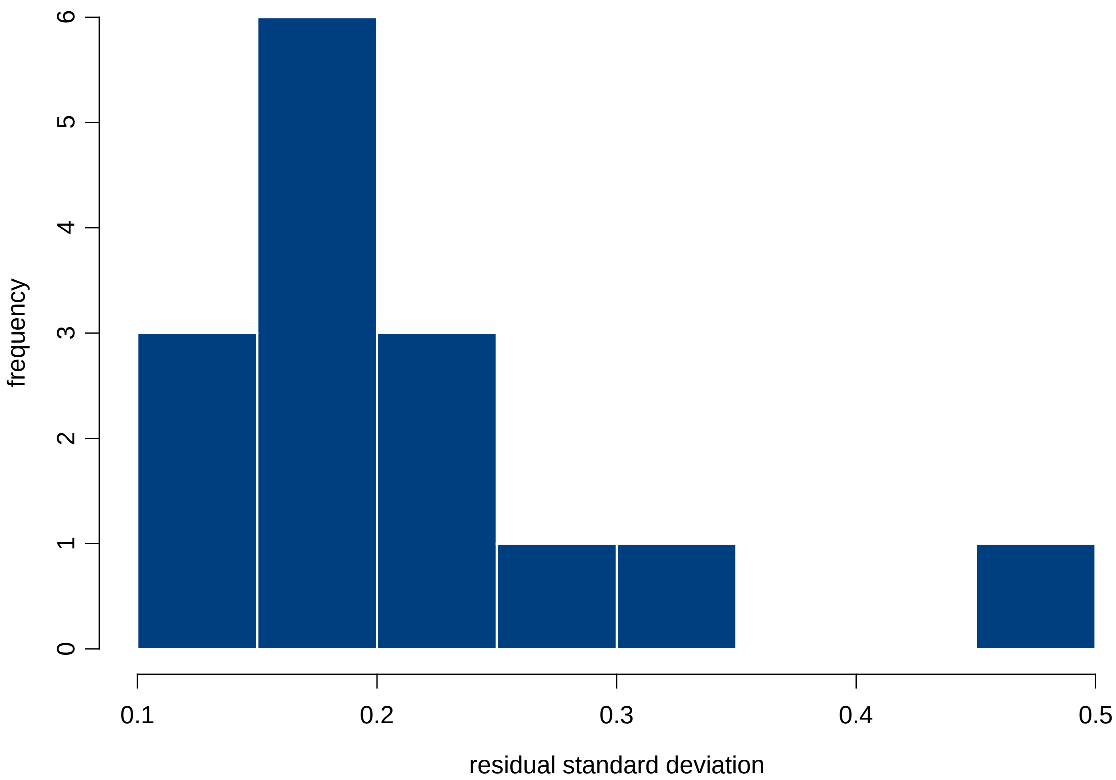

Figure 4. Histogram of within-district (or residual) standard deviations, σ_i for 15 Districts of the Habsburg Monarchy

Within-district (residual) standard deviation characterizes departures of (2) from the saturated (or "best possible") model (1). As it is tied to model error, it should be small compared to the real variation in height for a given year and district (i.e. to σ_{it}). To get some idea about the comparison, we plot $\sigma_{W,t} = \sqrt{\frac{\sum_{i=1}^{15} \hat{\sigma}_{it}^2}{15}}$ for each year t in Figure . These are average ("net biological") variations for each year. There is a clear downward trend showing tendency to within-district homogenization over time. $\sigma_{W,t}$'s can be compared to $\hat{\sigma}_i$'s from Figure . Notice that the residual errors after model (2) are certainly not larger than random sampling errors within districts, as described by $\sigma_{W,t}$.

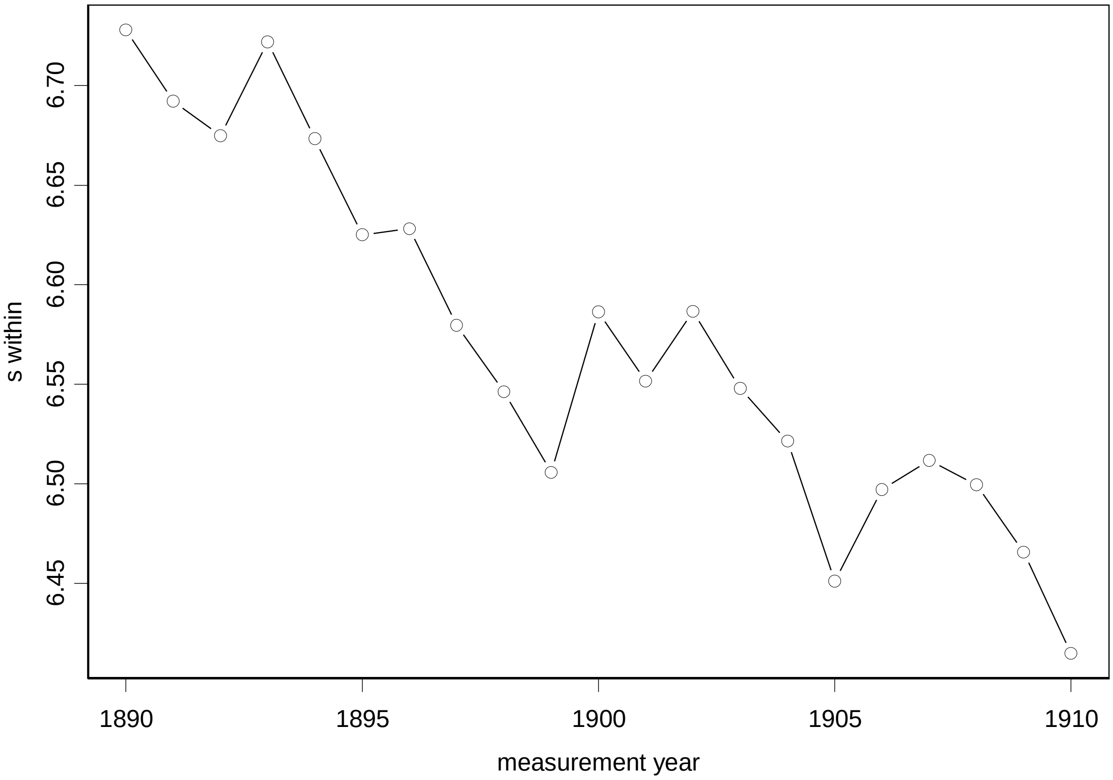

Figure 5. Within-district standard deviations ($\sigma_{W,t}$).

When one is interested in "convergence" or "spatial homogenization" of the Monarchy's districts in terms of height, it is possible to look at year by year between-district standard deviations - namely at $\sigma_{B,t} = \sqrt{\frac{\sum_{i=1}^{15}(\hat{\mu}_{it} - \hat{\mu}_{.t})^2}{14}}$ (where $\hat{\mu}_{.t} = \frac{\sum_{i=1}^{15}\hat{\mu}_{it}}{15}$) for each year *t*. There is a clear downward trend visible on this plot, suggesting that some between-district homogenization or "convergence of different districts to homogeneity" really took place during the 1890-1910 period.

Figure 6. Between-district standard deviations ($\sigma_{B,t}$).

In view of Figure , one possible criticism of the approach is that such an impression could, in principle, be implied entirely by "general homogenization" which would took place *within* each district that has nothing to do with their convergence to uniformity. To explore this possibility, we plot the ratio $\frac{\sigma_{B,t}}{\sigma_{W,t}}$ in Figure 7. One can see that not all apparent convergence could be explained by the overall homogenization taking place within all districts. Between-districts homogenization process is quicker than the within-district process. This can be seen roughly from Figure 7, and more formally e.g. from comparison of the slopes for linear approximations fitted for the data as shown in Figure 8. They are -0.0134 cm/year for within-district variability and -0.0227 cm/year for between-district variability. This finding provides a motivation for further investigation of the district convergence in a more structured and formal way.

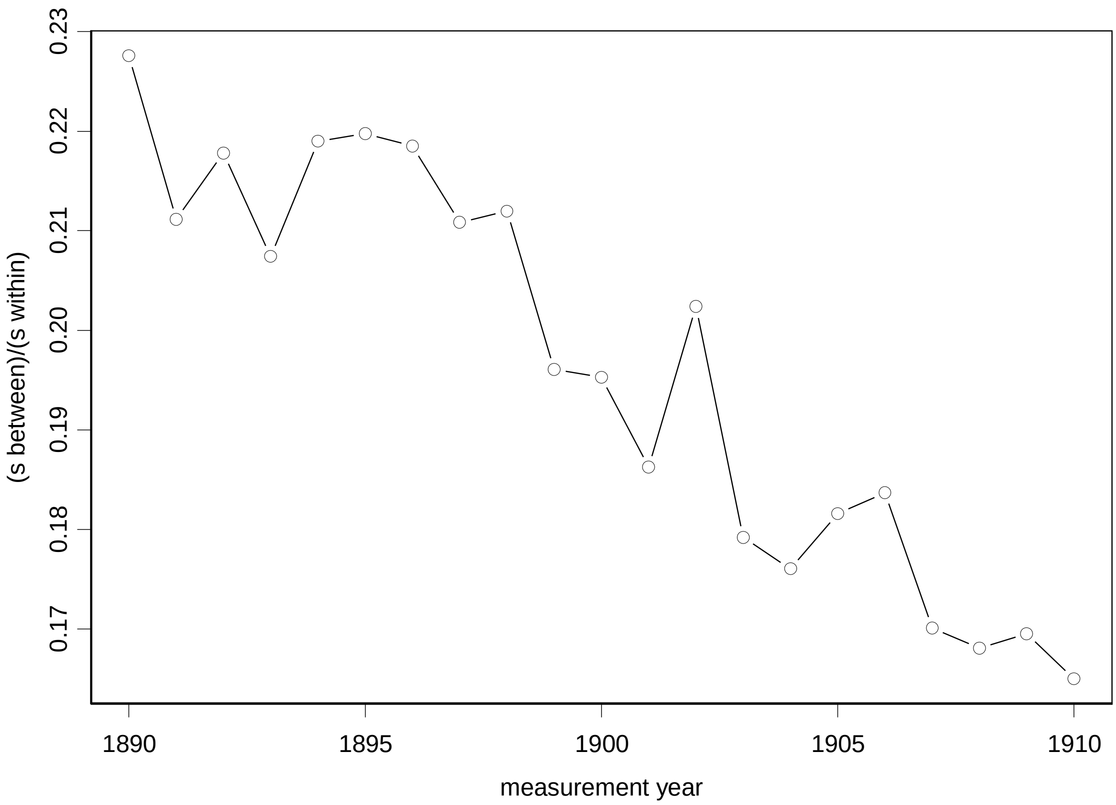

Figure 7. Between- to within-district standard deviations ratios.

Figure pools the information on intercepts and slopes (from model (2)) and shows district-specific estimates of the linear secular trends in height.

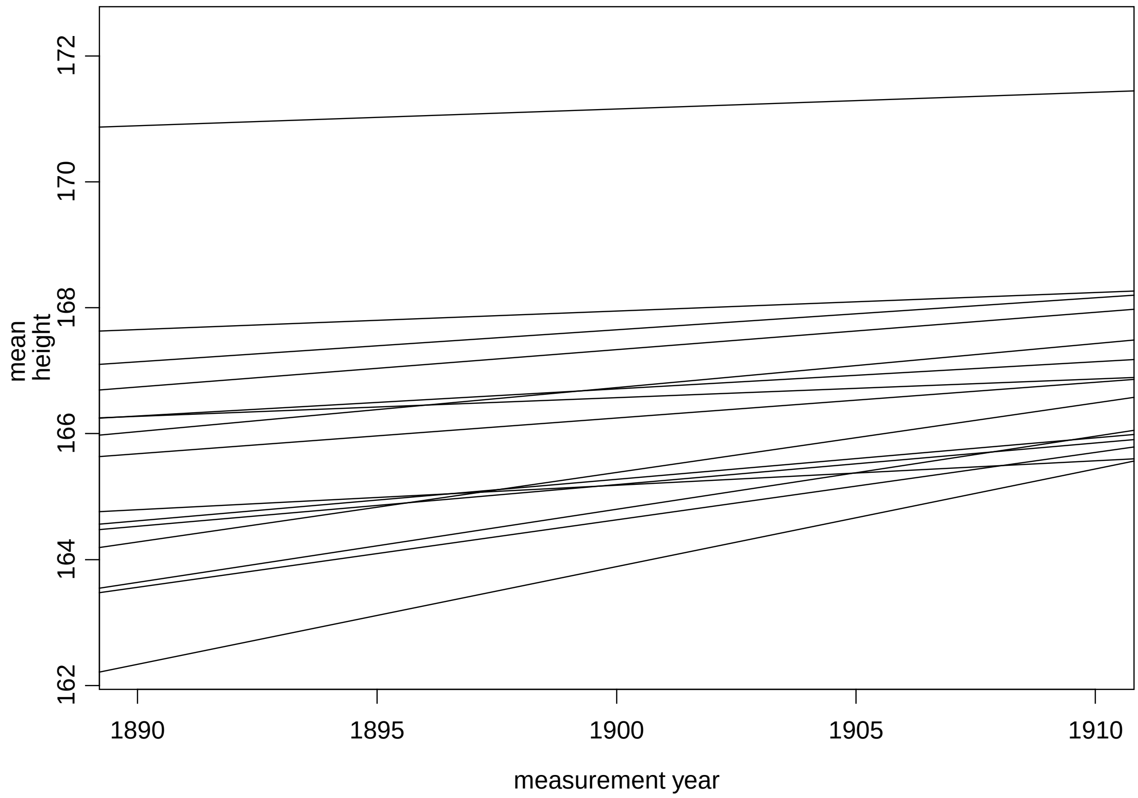

Figure 8 Estimated linear trends in heights (cm) for 15 districts of the Habsburg Monarchy.

Note that slope for initially lower-height districts tend to be higher than for those with greater mean height, i.e., secular trend depends (negatively) on initial conditions. Moreover, the spread over the cities decreases because the heights in the lower-height districts increase faster. This implies that the socio-economic and sanitary conditions were generally improving faster for youth and children in the periphery of the monarchy than in the core districts. The higher-height districts do not contribute to the equalization as much as the few initially lower height ones do. Thus, the slopes are dependent on the intercept (i.e., dependence of a simple summary of dynamics in the 1890-1910 period on the initial condition given by 1890 mean height), but the burden of the attenuation of initial inequality was not distributed equally nor symmetrically over the initial heights.

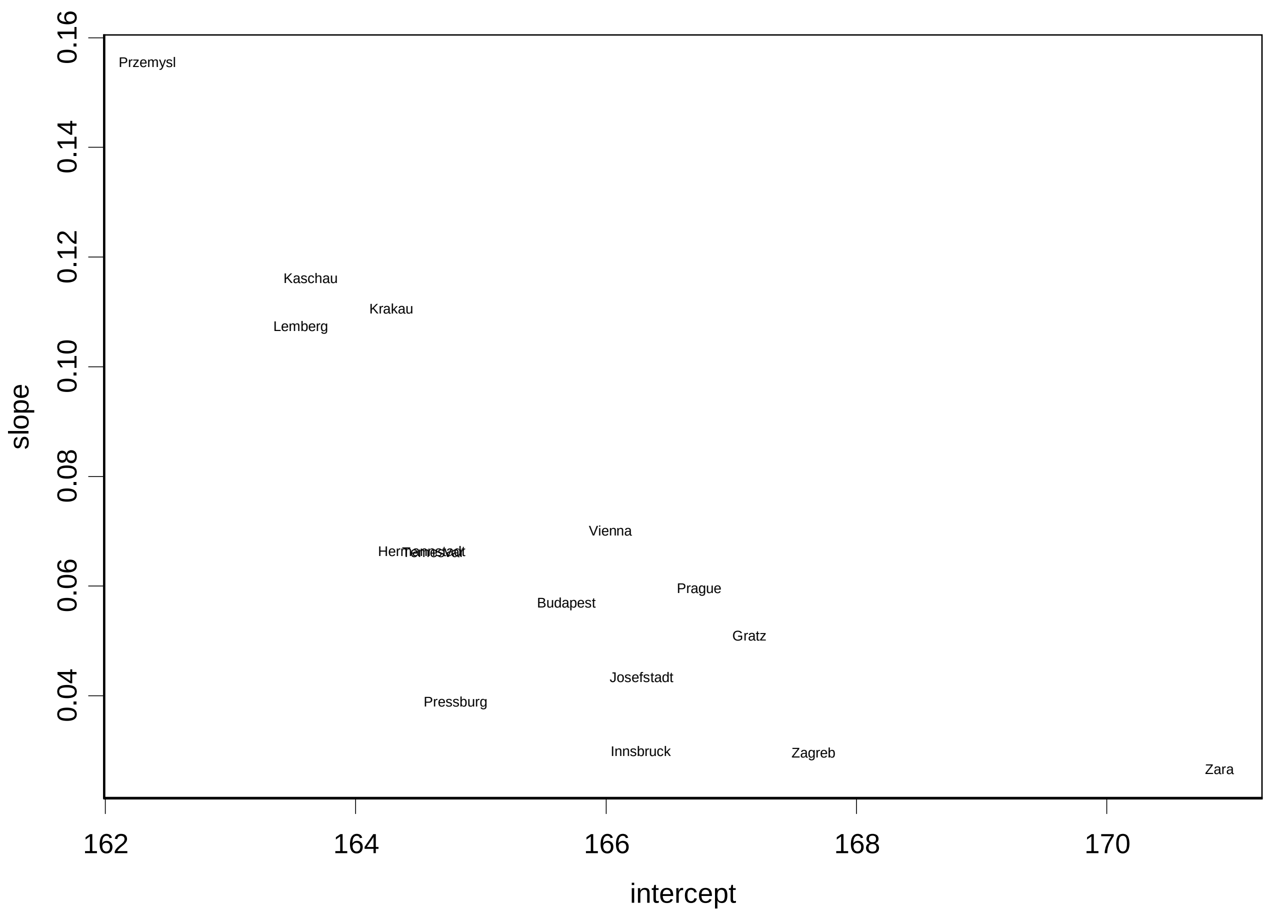

Figure 9. The estimated trend in height as a function of initial height in 15 districts of the Habsburg Monarchy.

Fitting a straight line through this plot yields: $slope_i = 2.451 - 0.014 intercept_i$, where the coefficient on $intercept_i$ (or on the initial condition when one considers the parameterization mentioned earlier) is significant (p=0.0004). As Zara (Dalmatia) is an outlier with small if any progress realized over the 1890-1910 period, we refitted the line to the data from all districts except for Zara and obtained: $slope_i = 3.481 - 0.021 intercept_i$ (p=0.0001). Qualitatively, nothing substantial changes, except for the fact that the dependence of mean secular district-specific velocity on initial condition strengthens even more. Generally, the negative relationship acts as a balancing mechanism similar to feedback.

One might be interested in constructing some “overall” estimates, aggregating over the districts. To that aim, we introduce the following linear mixed model (or random coefficient model)[5]:

$$Y_{itk} = \beta_0 + \beta_1(t-1890) + b_{0i} + b_{1i}(t-1890) + \varepsilon_{itk} \qquad (3)$$

where Y_{itk} is the height of a man, randomly selected in district *i* and year *t*. β_0, β_1 are fixed effects corresponding to Habsburg male population-averaged (or more specifically, spatially-averaged) regression coefficients. b_{0i}, b_{1i} are district-specific random coefficients, assumed to have $\begin{pmatrix} b_{0i} \\ b_{1i} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{01} & \sigma_1^2 \end{pmatrix}\right)$ distribution. Further, we assume $\varepsilon_{it} \sim N(0, \sigma^2)$, independently across *i* and *t* and independently from b_{0i}, b_{1i}.

Since Zara seems to behave quite differently from other districts, we fitted the model without Zara (although including it in the analysis did not change the results substantially) and got spatially averaged intercept estimate $\hat{\beta}_0 = 165.243$ with 95% confidence interval (164.439;166.046) and spatially averaged slope estimate $\hat{\beta}_1 = 0.072$ with 95% confidence interval (0.053;0.092). In other words, there is a strong (positive) linear secular trend on average, as expected. For every cm below the average height of 165.243 cm, height at the district level would have increased by 0.072 cm on average (averaging across space and time). Standard deviations corresponding to individual variance components are estimated as: $\hat{\sigma}_0 = 1.525$, $\hat{\sigma}_1 = 0.036$, $\hat{\sigma} = 0.203$ (all of these were fitted by REML[6]). Hence, there is non-negligible variability in both initial conditions and average secular velocities. This implies that the standard deviation of initial heights around their mean of 165.243 cm is 1.525 cm and that the standard deviation of the slope coefficient about their mean is 0.036 cm (Where mean is meant for spatial averaging (across districts) here). Moreover, one obtains estimate of $\hat{\rho}_{0,1} = corr(b_{0i}, b_{1i}) = \text{-}0.858$ indicating very strong negative relationship between initial condition and slope by district. This corroborates the findings of Figure by alternative and more formal means.

Spatial view

We explore how the random coefficients of model (3) (or equivalently the random effects b_{0i}, b_{1i}) behave in space (among districts). One might consider that for such a large area as the Habsburg Monarchy the distance of individual districts from the center (Vienna) should be important for economic development (areas closer to the center might evolve more quickly, while more distant areas might be slower in development). Note that to define the distances in ideal way is a very difficult task for two reasons: i) we are dealing with districts of relatively large areas (not points), ii) Euclidean distance might not be exactly what is needed (distance along main terrestrial communications used at the time would be more relevant). Nevertheless, we use the Euclidean distances among cities listed as military (not geometric) centers as proxies. Then we can explore separately how district specific random effects b_{0i} and b_{1i} (i.e. departures of a particular district's initial condition from spatially-averaged initial condition and departures of a particular district's secular velocity from spatially-averaged velocity, respectively). The random effects (together with their locally quadratic loess smooth[7]) are shown Figure and Figure . b_{0i} and b_{1i} are from Eq 3.

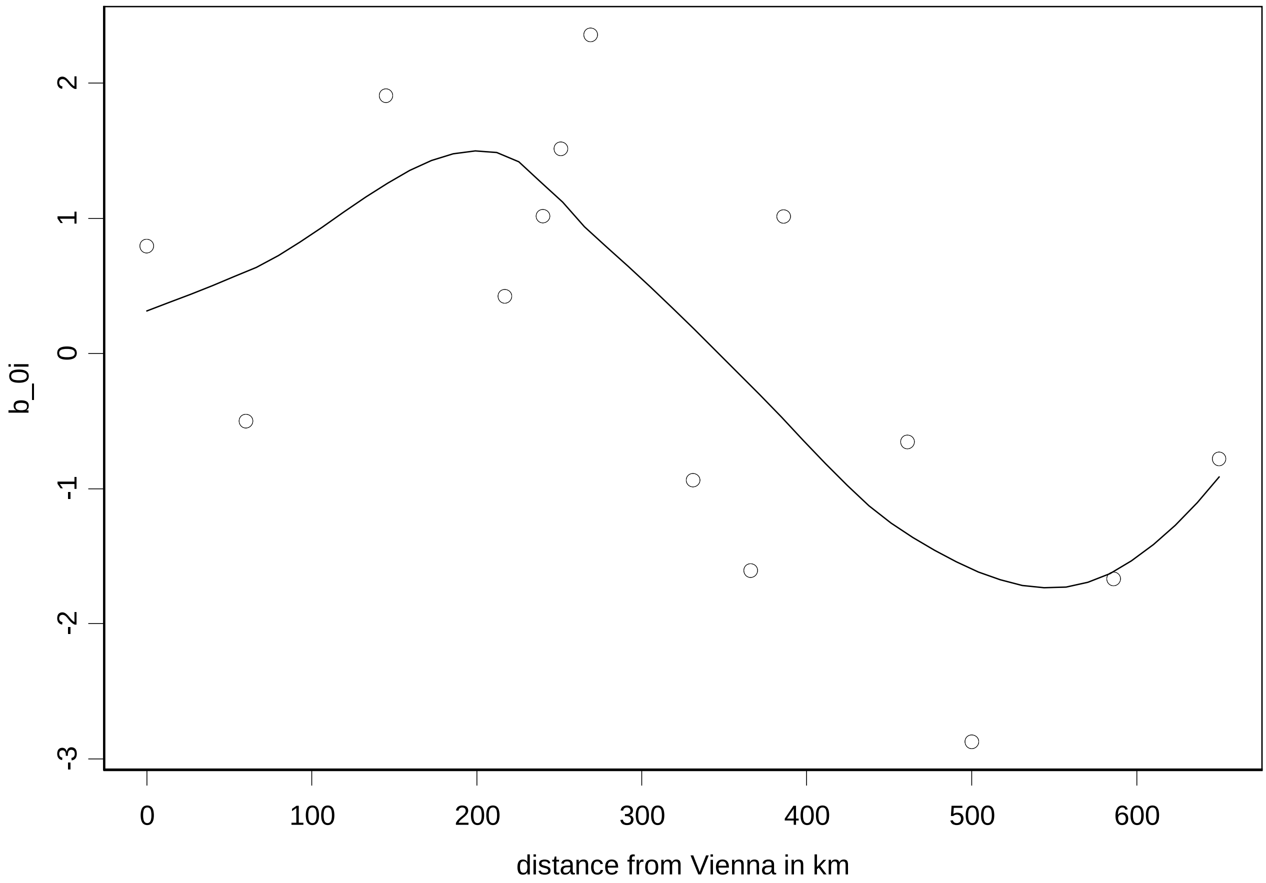

Figure 10. Dependency of b_{0i} (district level heights from mean heights in 1890 in cm) on distance from Vienna.

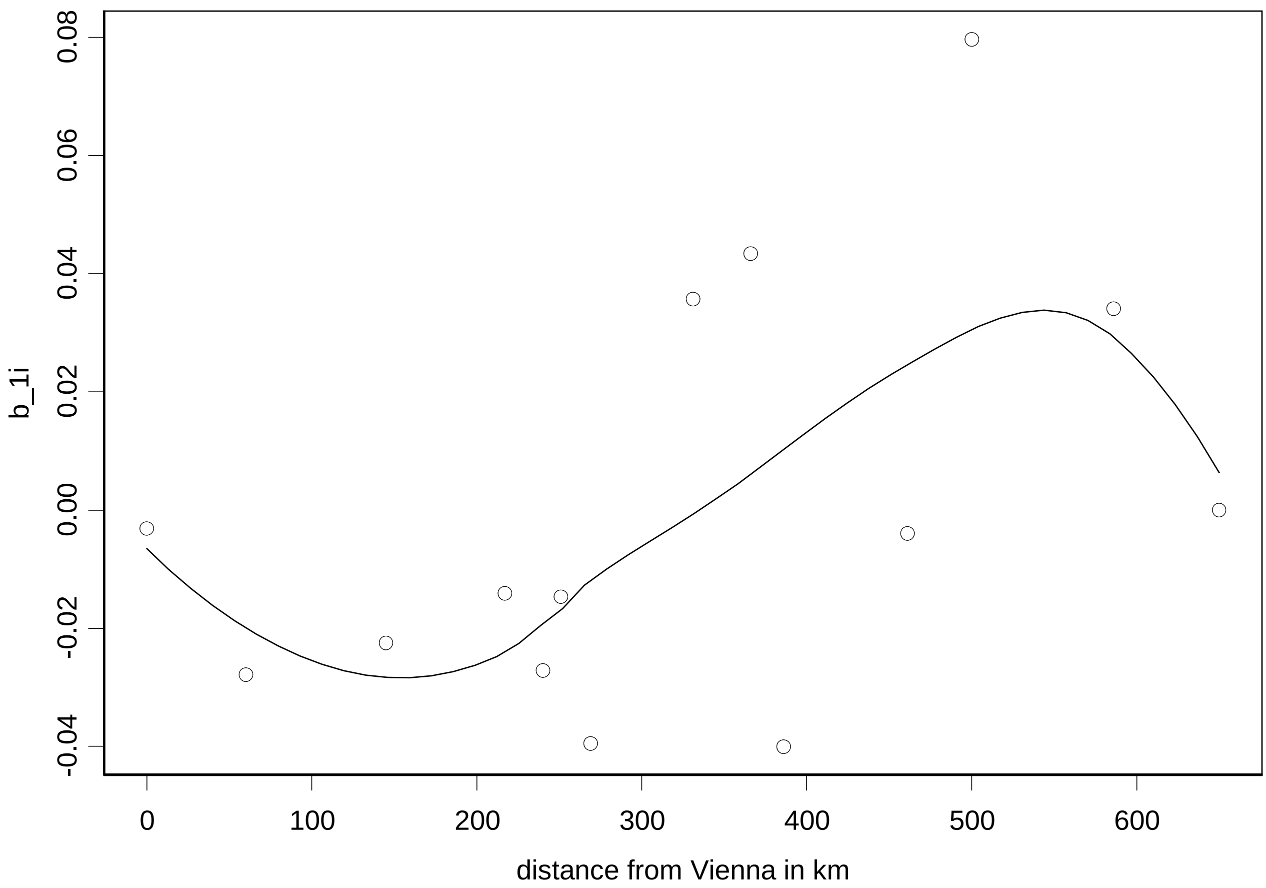

Figure 11 Dependency of b_{1i} (in cm/year) (district level deviation from average secular trend) on distance from Vienna.

While ideally, one might expect monotone relationship between b_{0i} and b_{1i} ad distance from Vienna, the situation is apparently more complicated (probably due to other local influences upon economic and human stature development than just straight-line distances from the monarchy's center). It seems that both in intercept and slope, there is a minimum and maximum. Although they are not very pronounced, their possible interpretation is interesting. Figure 10 suggests that initial conditions are better close to Vienna, while Figure 11 suggests that the fastest improvement occurred (in 1890-1910 period) in regions far away from the center. This impression can be verified formally by extending model (3) to:

$$
\begin{aligned}
Y_{itk} = {} & \beta_0 + \beta_1(t-1890) + \\
& \gamma_{01} d_i + \gamma_{02} d_i^2 + \gamma_{03} d_i^3 + \\
& \gamma_{11} d_i (t-1890) + \gamma_{12} d_i^2 (t-1890) + \gamma_{13} d_i^3 (t-1890) + \\
& b_{0i} + b_{1i}(t-1890) + \varepsilon_{itk}
\end{aligned}
\tag{4}
$$

where the previously defined symbols and assumptions on random effects remain the same as in (3). d_i is the distance between the military center of i-th district from Vienna (in km). Newly introduced γ coefficients describe modification of spatially-averaged intercepts and slopes with distance from Vienna. This modification is modeled with a third order polynomial (to capture two local extremes seen in Figure and Figure), separately for intercept and for slope. Likelihood ratio test of (4) versus (3) gives p-value=0.0648. Therefore, equation 4 is preferred at the 7 % significance level. The cubic coefficients for both intercept (γ_{03}) and slope (γ_{13}) modifications are significant (p-values 0.0468 and 0.0473, respectively). One can then solve the quadratic equation arising from the condition for local extremes with the estimated coefficients ($\hat{\gamma}$, MLEs) to get the locations of local minimum 547 km, local maximum 145 km for intercept modification and local minimum 133 km, local maximum 518 km for slope, quite similarly to the impression obtained from Figure and Figure .

Another question from the spatial context can be: is it true in general that districts that are closer (in crude geographical sense discussed above and) are also more similar in height even after we correct for the distance to Vienna in the model (4) sense? To this end, we need to consider not only distances from a given district to Vienna as before, but we need to take all pairs of (different) districts (there are $\binom{14}{2} = 91$ of them if we exclude Zara from the analysis as an outlier). Note also that for these considerations we need to clean the data for the trend in time before taking spatial correlations (or correlations among districts). We do it by computing spatial correlations for each district pair from residuals after model (4). This checks if there is any spatial correlation left in addition to the spatial structure implied by the simple „distance from the center only" structure. For technical reasons, we work with Pearson residuals (i.e. residuals divided by their model-based standard errors to correct for the fact that the residuals, unlike the true ε's are not iid). Figure 12 shows the correlations as a function of distance between districts (dots). This is essentially the spatial correlogram[8]. To see a possible pattern more clearly, loess nonparametric smooth is superimposed.

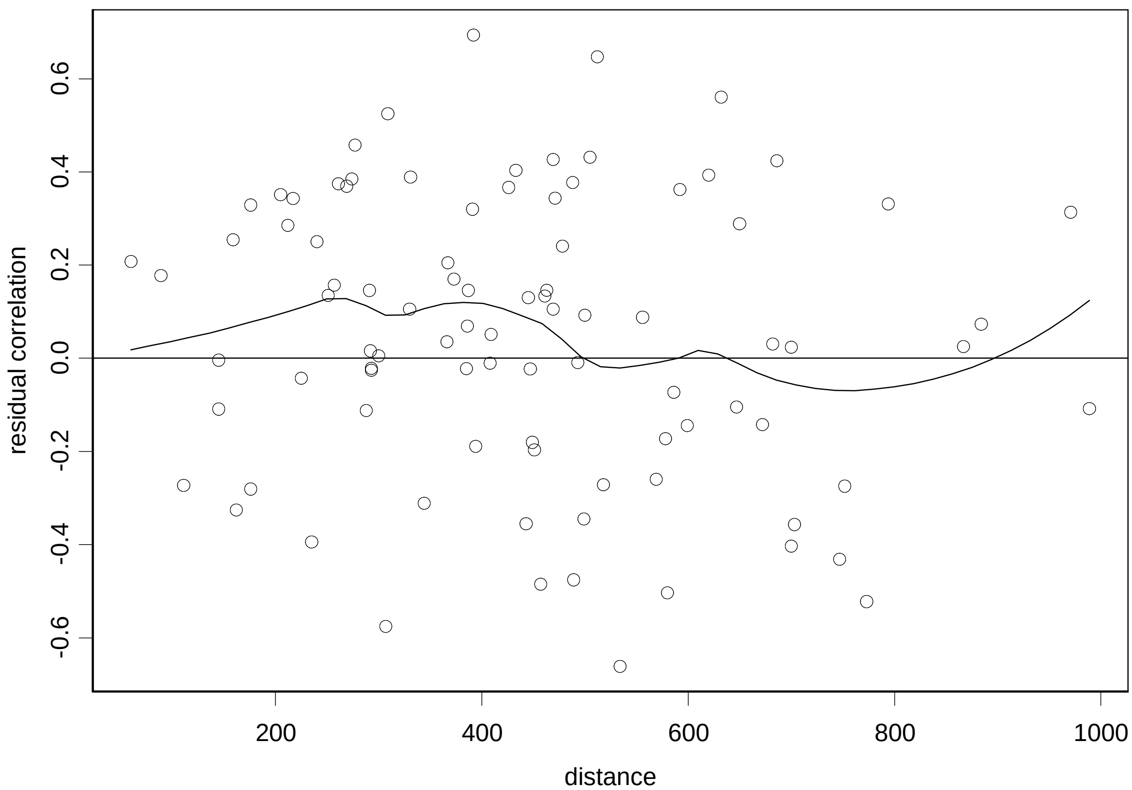

Figure 32 Residual correlations of heights as a function of distance between districts

No strong pattern consistent with "anything that is closer tends to be more correlated" hypothesis is discernable here. In fact, the correlations are typically rather small, indicating no substantial problems in model (4), as a byproduct. It seems that spatial dependence: i) concentrates into the structural correlation among random effects and not to the residuals, ii) instead of general spatial dependence, strong relationship to Vienna as the center is preferred.

Test of the linearity of the trajectories

A question remains, whether there are systematic features beyond the linearity of growth assumed above. Many statistical tests can be constructed to "test for nonlinearity". We will do two LOF (Lack Of Fit) tests to check the quality of the fit: i) compare linear (2) with quadratic model in time (as the most simple nonlinear model), ii) compare linear trend model (2) with smooth trend model (fitted by nonparametric loess smoother in default version with span=0.5 and local linearity, allowing for globally nonlinear, but smooth shape). They were applied for each district separately (using weighted regression, in a manner similar to that used in model (2), i.e. with inverse square SEM weighting). Note that ii) compares the same null, i.e. linear model to much broader set of alternatives than i), hence it is more general but less powerful. Resulting p-values are listed in Table 2:

Vienna	0.6781	0.2746
Prague	0.2967	0.0477
Gratz	**0.0016**	0.0001
Innsbruck	0.0327	0.0002
Zara	0.0144	0.0393
Lemberg	0.0105	0.0084
Budapest	0.0459	0.0255
Hermannstadt	0.0102	0.0013
Zagreb	0.0403	0.0008
Krakau	0.7224	0.3746
Josefstadt	0.0318	0.0015
Pressburg	**0.0001**	0.0009
Kaschau	**0.0000**	0.0000
Temesvar	0.2363	0.0049
Przemysl	0.5606	0.5276

Table 2 p-values for LOF tests

At the 0.01 level, i) finds 3 substantially nonlinear trajectories (Gratz, Pressburg and Kaschau), while ii) there are some departures from linearity for 9 districts, but they are generally not of simple shape (like exponential or polynomial). Nevertheless, they should be investigated. One way to extend the linear model (2) is to make it more flexible as follows:

$$Y_{itk} = \beta_{0i} + \beta_{1i}(t-1890) + f_i(t-1890) + \varepsilon_{itk} \qquad (5)$$

where Y_{itk} is the height of a man randomly selected in district i and year t (). The average slopes and intercepts for 1890-1910 are β_{0i} and β_{1i} (as estimated from (2)), while the district-specific curves $f_i(,)$ are systematic departures from linearity. $f_i(,)$ are assumed to be smooth. Effectively, these are found by the locally linear loess fit with span = 0.5. As before, $\varepsilon_{it} \sim N(0,\sigma_i^2)$ (independently across i and t) is assumed. Note that (5) is just a reparametrization of the model used for testing LOF in ii) above and that (5) explores behavior of residuals in quite explicit way. In this sense, the analysis here extends (and does not contradict) the linear analysis based on model (2). Ideally, residuals should have constant (zero) mean. Figure shows $f_i(,)$ curves for all 15 districts. Due to the fact that we are working with residuals from the linear trend, the horizontal line at 0 in the following graph corresponds exactly to the straight line trajectory fitted under model (2).

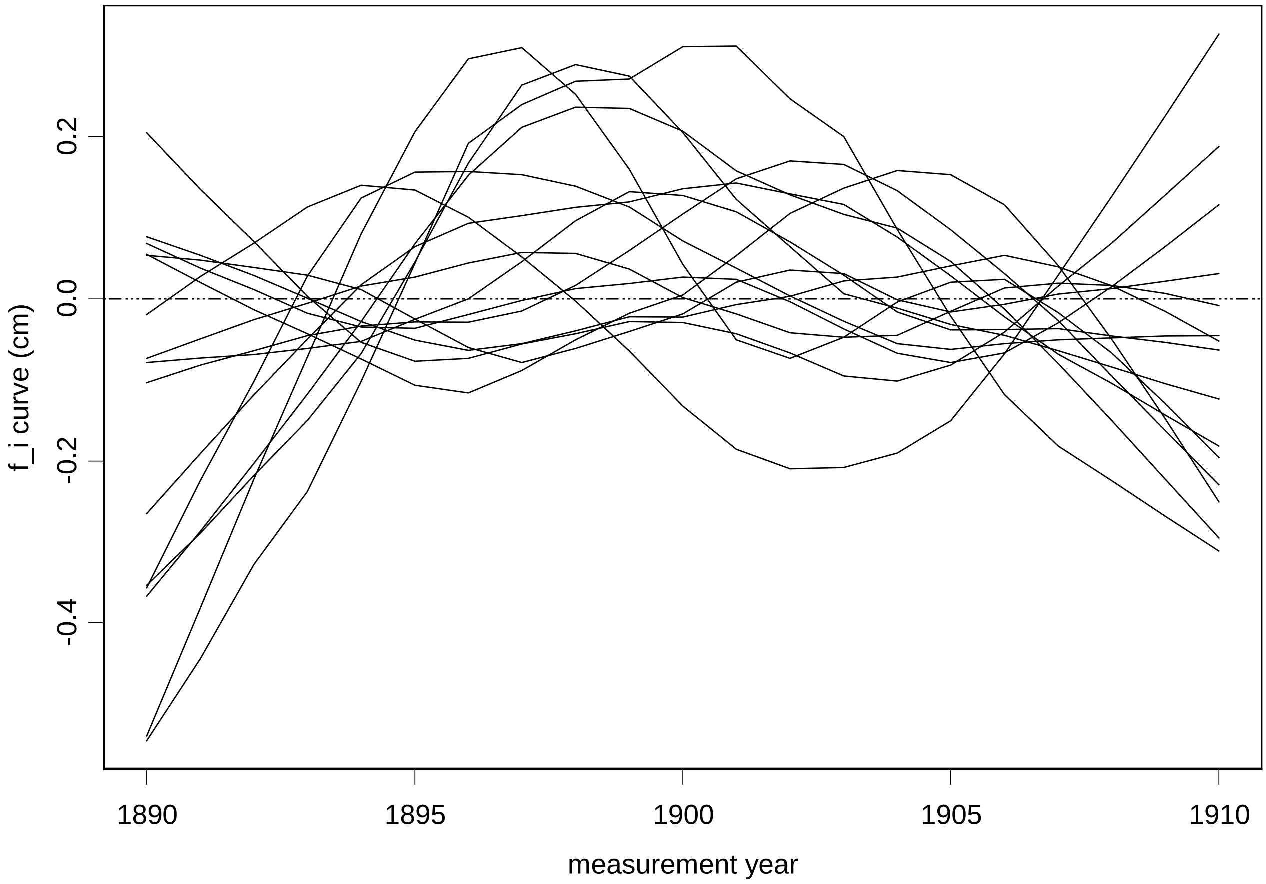

Figure 13. Smooth trends in residuals

Notice that while there seem to be some systematic patterns of departures from linearity, some of which even repeat over several cities, they are not simple. In particular, they are not monotonic. They might suggest either periodicity or slight hints of concavity in original tracks (with one or two exceptions). If this is hard to see from the residuals, we provide reparametrization of model (5), with

$$g_i(t) = [\beta_{0i} + \beta_{1i}(t-1890) + f_i(t-1890)] \qquad (6)$$

and plot the district-specific $g_i(.)$'s in Figure 4[Marek Bra2]. The districts are divided into two groups, the economically more advanced areas and the peripheral areas. Heights in the advanced areas were all above average at the beginning of the period, and remained so, in the main, throughout, while heights in the peripheral areas were below average except for Zagreb and the outlier Zara.

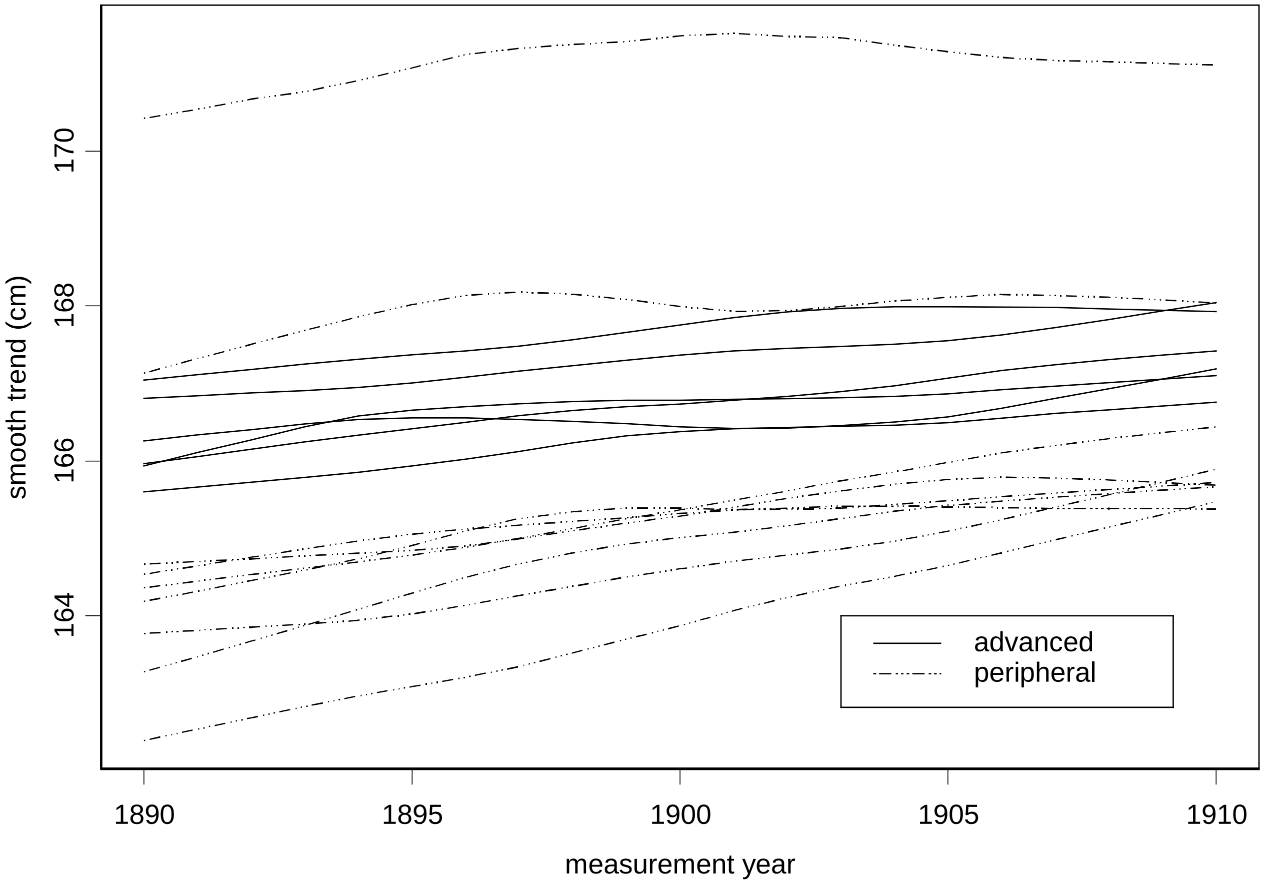

Figure 44. Reparametrization (8) of model (7) . Advanced areas: 1=Vienna, 2=Prague, 3=Gratz, 4=Innsbruck, 7=Budapest, 11=Josefstadt (Bohemia). Peripheral areas: 5= Zara, 6=Lemberg (Galicia), 8=Hermannstadt (Transylvania), 9=Zagreb (Croatia), 10=Krakau (Galicia), 12=Pressburg (Slovakia), 13=Kaschau (Slovakia), 14=Temesvar (Transylvania), 15=Przemysl (Galicia).

Dependency on economic variables

The analyses performed so far use time and distance as the two independent explanatory variables. It might be also interesting to take into account economic information related to secular height changes. This poses two problems. The first problem is inherently related to the way biological and economic processes are related: what kinds of variables are relevant (GDP, industrial output, agricultural output, food prices, cost of living index). These are generally not available in good quality for historical periods. Moreover, the second problem is that even if some of the indicators are available at aggregate level for the Monarchy, they are not at our disposal at the disaggregated level by districts. Even though it would be ideal to estimate the relationship between height and selected socio-economic indicators for each district separately (essentially by doing stratified regressions or via a mixed model, assuming some degree of similarity across districts), in the absence of such detailed information, one can at least a regression of height on socio-economic indicators at the aggregate level in order to obtain a relationship between individual district heights and the average indicator. Departures from the relationship to the average indicator can then be analyzed and interpreted in more powerful ways than analyses without the indicator.

The first problem is harder, though, because theoretically there are many potentially relevant indicators. Even if we concentrate on just three variables for the Monarchy: agricultural output, industrial output, population, there are many possible forms in which to take them into a model. This multitude pertains not only to the functional forms of the indicators (linear, higher polynomial, log, etc.), but also to the fact that it is not clear in which lag to take the explanatory variable. On one hand, it is clear that even if we take as granted that there is a direct relationship between the given economic indicator and height (say at 21 years, as relevant here), it is to be expected on basic biological grounds that it takes some time before a change in the indicator translates into a height change. Obviously, one would not expect a direct relationship between height measured in year t and economic indicator of year t. In other words, one would expect some lag l and relate height in t to the indicator in $t-l$. Note that another simple solution $l=21$ as the relationship to the economic situation at the approximate time of birth (precision within a year is all what we consider) would not solve the problem either. In fact, one would expect some intermediate situation in which l would be related to the time when physical growth is both rapid and sensitive to outside conditions. It is clear that such a question is hard to solve by statistical means only. Biological insight should be used (using e.g. biological and medical interpretation of the mean growth curve in the 1890-1910 period as well as other rather complex a priori biological information), instead. We do not have such information at our disposal, so we attempt a rather approximate and pragmatic approach. It is based on fitting the following model with a single given indicator for different lag *l*, then considering the evaluation of the likelihood function at the MLE estimate (which can be compared across lags and across indicators).[9]

$$Y_{itk} = \beta_0 + \beta_1 I_{t-l} + b_{0i} + b_{1i} I_{t-l} + \varepsilon_{itk} \qquad (7)$$

where assumptions about errors and random effects is the same as in model (3). Y_{itk} is the height in *i*-th district in year *t*, I_{t-l} is the value of a given indicator at time $t-l$ (in other words, *l* is the lag i.=1...21). For the indicators, we choose: 1) agricultural output per capita, 2) industrial output per capita, 3) total population. All three of them are for the Monarchy as a whole. The fit was performed after excluding Zara.

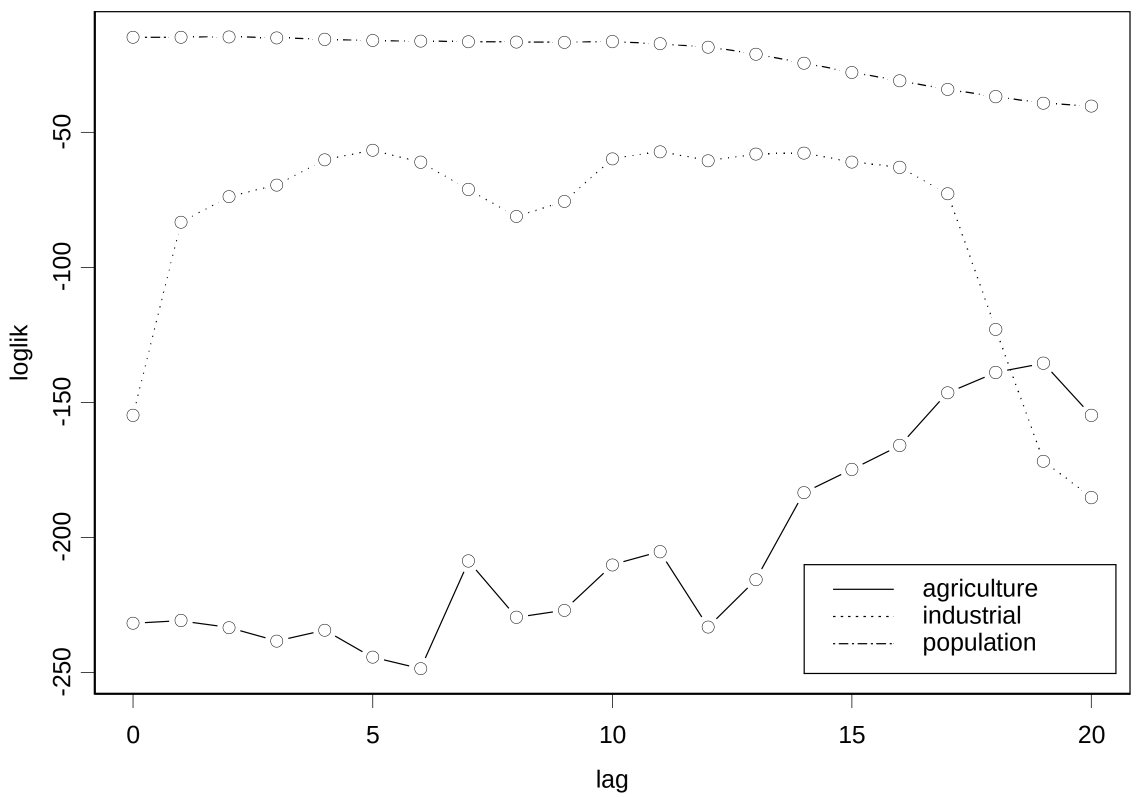

Figure 15. Loglikehood values for various lags of Eq. 7.

The loglikelihood values for 63 regressions (21*3=63) are shown in Figure 15. Different points correspond to different models (with different lag and/or indicator) being fitted, nevertheless all of them are of the type given by Eq. (7). Each regression is given by the combination of the indicator type and various lags (from 1 to 21).[10]

It is apparent that the various indicators have different explanatory power (as judged by the loglikelihood value at MLE. Paradoxically, agricultural output seems to have generally smallest explanatory power and population the highest. Based on assumed direct food -> stature relationship, one would expect an exactly opposite ordering! Apart from possible sophisticated ecological arguments that might promote the role of crowding in population regulation (usually falling under the "density dependence" keyword in theoretical ecology), the explanation might be much less subtle. Possibly, there is a substantial imprecision in the

Monarchy's statistics, possibly resulting in some attenuation to regression coefficients. If that is so, one would expect that population is given with largest precision, agriculture output (necessarily based on counts or estimates of many small quantities hard to measure) with the smallest. Certainly, one should consider this possibility before a hasty conclusion. Next, despite generally small values of the loglikelihood for agricultural output, shape of the curve across lags is probably most plausible. Lag giving maximum loglikelihood is 19 (corresponding to 2 years of age). This might be quite close to what one would generally expect. On the other hand, lags giving maximum loglikelihoods are 5 for industrial output and 2 for population. These correspond to 16 and 19 years of age and therefore seem to be rather late in adolescence to be too much influenced by environmental conditions directly. Nevertheless, for the purpose of fitting a model with all three indicators simultaneously, we took the maximum lags as shown yielding the model:[11]

$$Y_{itk} = \beta_0 + \beta_1(t-1890) + \beta_2 A_{t-19} + \beta_3 D_{t-5} + \beta_4 P_{t-2} + b_{0i} + b_{1i}(t-1890) + \varepsilon_{itk} \qquad (8)$$

where the assumptions on errors and random effects is the same as before. Y_{itk} is district-level height, A is the Monarchy's agricultural output per capita, D is the Monarchy's industrial output per capita and P is the Monarchy's population. Further, we assume

$\varepsilon_{it} \sim N(0, \sigma_i^2)$, (independently across *i* and *t*) and independently from b_{0i}, b_{1i}.

This is a model similar to (3) (fixed effects structure), except that it adds more explanatory variables, with lags selected in the previous univariate optimizations. Presumably, such a model might be more powerful in elucidating inter-district differences and their changes through time. Indeed, likelihood ratio test indicates that (8) is substantially better than (3), with p-value=0.0030. On the other hand, tests for individual coefficients give p-values: 0.0049 for β_1, 0.9347 for β_2, 0.5222 for β_3, 0.0147 for β_4. This suggests that there is a strong relationship of height to time and population lagged by 2 years. Values of the MLE estimates for the β_1, β_2, β_3, β_4 coefficients are 0.4063, 0.0000, -0.0019, -0.0008 respectively. Note the negative relationships with industrial output and total population. Biologically, this does not seem to be extremely elucidating model. Nevertheless, its explanatory power is substantial.[12]

A model working with just one lag might serve as a good first approximation, but is not entirely feasible on biological grounds. Instead, a model with simultaneous relationship to several lags (or a distributed lag model) should be more plausible. This is somewhat similar to the problem of (time-invariant) filter identification, frequently encountered in various applied fields dealing with dynamic relationships, like engineering and economics. However, its fit to real data poses quite a challenging task. This is mainly due to the fact that due to smoothness and generally increasing character of a typical explanatory indicator, different lags tend to be substantially correlated. One solution to the problem is to use Almon's polynomial lag model (Johnson 1984). Motivated by rough biological considerations and by the shape of the agriculture output lag optimization curve in Figure 14, we will use the Almon lags (in a broader mixed effect model) for agricultural output only.

$$Y_{it} = \beta_0 + \sum_{l=0}^{20} \gamma_l A_{t-l} + b_{0i} + \varepsilon_{it} \tag{9}$$

where $b_{0i} \sim N(0, \sigma_0^2)$, (independently across i and t) and independently from b_{0i}. A_{t-l} is the Monarchy's agricultural output per capita, lagged by *l* years (corresponding to influence at 21-*l* years of age). γ_l is the coefficient on the *l*-th lag. Almon's K-th order polynomial lag model, is given as $\gamma_l = \sum_{k=0}^{K} \alpha_k l^k$. Therefore, model (9) does not have 21 fixed coefficients as it appears on the surface, but just K+1, hence achieving much parsimony (through the polynomial restrictions imposed on the lag coefficients). We choose $K = 5$ to accommodate up to 4 possible local extremes in the γ_l versus l curve (roughly suggested by the lag optimization plot).

After fitting the random coefficient model (9) (in fact, it is the random intercept model) by MLE, we get $\hat{\sigma}_0 = 1.184$ and $\hat{\sigma} = 0.290$, roughly comparable to the corresponding figures (1.469 and 0.203) of model (3), whose random coefficient structure is richer and hence it has smaller $\hat{\sigma}$. The Almon-like α coefficients can be transformed into the γ coefficients that can be plotted and interpreted. Figure 15 shows them as a function of lag index.

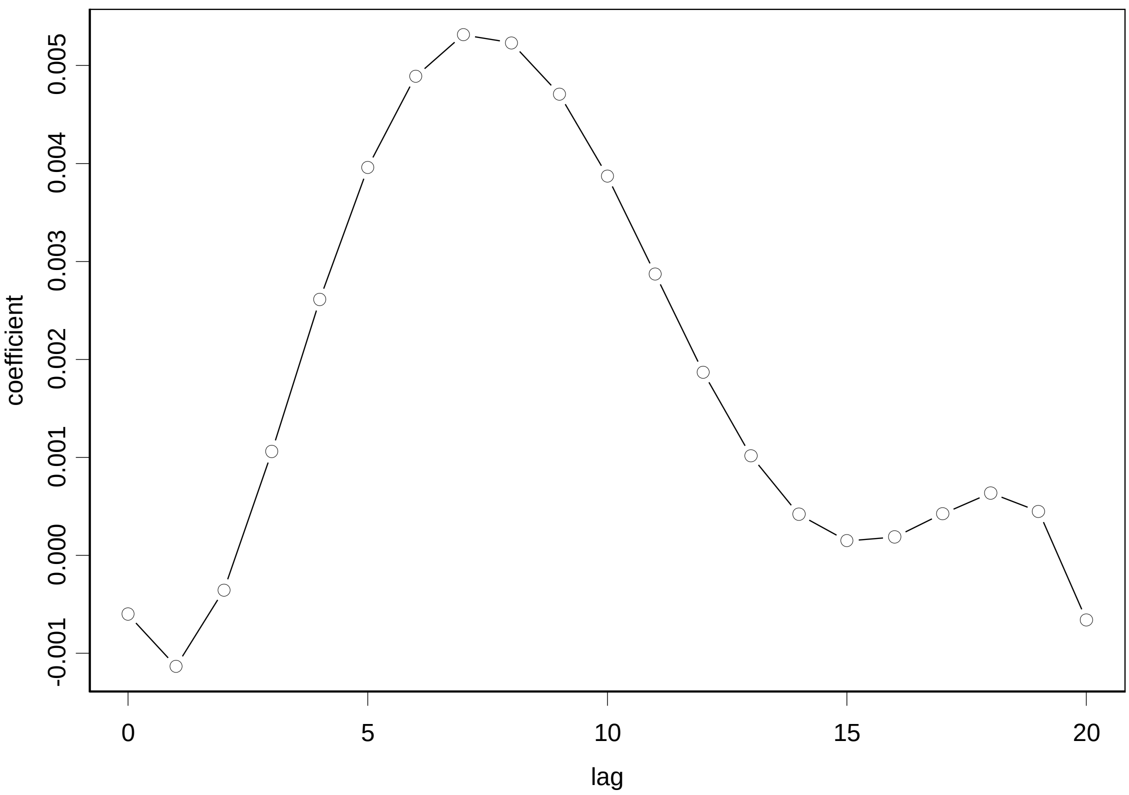

Figure 16. Lag structure obtained via Almon approach for agricultural output.

The pattern looks quite different from that based on single lag maximization in Figure 14. Although it has a local maximum close to that found by a single lag maximization (lag 18, corresponding to age 3), it turns out to be a local maximum. Global one occurs at lag 7 (age 14) which is close to the the time of the beginning of the adolescent growth spurt. The filter structure seems to have quite sensible biological explanation: mature height (at 21) tended to be influenced (in 1890-1910 period) by the environmental conditions related to food availability (approximated by the per capita agricultural output) mostly around puberty and then to a smaller extent by early childhood conditions (their influence on adult height might be diminished by compensatory growth occurring later in childhood when conditions improve).

Conclusion

There is a consensus among economic historians of the Habsburg Monarchy that there were great discrepancies in all relevant indicators of development between the developed core and the peripheral areas (Good, 1984; Komlos 1983; Schulze 2000). The height indicator we are using as a proxy variable for the biological standard of living points in the same direction. Discrepancies in height were large among the military recruits measured at age 21 in 1890. The Polish, Slovak and Romanian areas had about half the GDP per capita of Austria. They also had the shortest populations. At the district level there was a 3.7 cm gap between the districts of Vienna (in Austria) and Przemysl (in Poland) which reflects a very large discrepancy in living standards between the peripheral areas and the more developed parts even if we disregard the values for Zagreb and Zara (Croatia) for being outliers (Table 1). The shortest recruits were found in the Polish, Transylvanian and Eastern Slovak districts. The recruits here were about as tall as the Mediterranean populations at the time, whereas the height of Austrians and Czechs were closer to Western-European levels. We were also able to confirm that generally speaking the farther were the districts where the recruits lived from the center of the Monarchy (Vienna), the shorter were the recruits by approximately 0.46 cm per 100km distance from Vienna in 1890 (Figure 10).

There was a tendency of heights to increase as well as to converge in this period. Hence, the anthropometric evidence indicates that there were economic advances in the Monarchy during the last few decades of its existence as conventional indicators have suggested. Yet, these advances locate the Monarchy in the middle-range of the European experience, further substantiating inferences drawn on the basis of conventional indicators. Heights increased slower than in Norway, Sweden or in the Netherlands, for instance, although they started at a higher level than did Austria.

Research on economic convergence among the various regions of the Monarchy has been somewhat limited on account of the paucity of regional evidence. Yet, the degree to which welfare converged across the disparate regions of the Monarchy is crucial to the judgment on its protracted conflict of nationalities before its ultimate political demise. Evidence on height is also used to consider spatial variability of biological living standards and its changes during the years 1890-1910.[13] We use two measures of spatial convergence: the standard deviation and the secular trend over time. We find that the shorter was the population in 1890 the faster it grew. Hence, there was convergence between the peripheral and the core areas of the monarchy. The difference between the trend in the height of the Polish district of Przemysl (1.6 millimeter per annum) and the Viennese trend (0.7 millimeter) was about 0.9 millimeter per annum (or 0.9 cm per decade). At that rate it would have taken men in the Polish district of Przemysl 43 years to catch up to Viennese men. In other words, convergence proceeded at a relatively slow rate. The standard deviation of heights within the districts themselves also decreased over time, implying that inequality declined also within districts, not only across them.

But the convergence among the core districts (located in today's Austria, Czech Republic, and Hungary) was minimal or non-existent, whereas the convergence among the peripheral districts (located in today's Poland, Romania, and Slovakia) was more rapid. Hence, spatial convergence took place exclusively within the peripheral areas, and between the peripheral regions themselves and the more developed ones. The pattern is somewhat reminiscent of modern findings on convergence clubs in the global economy: while there is convergence in income within OECD countries, there is little evidence of convergence of income to OECD levels from underdeveloped economies (Baumol, 1986; Galor, 1996). However, the Habsburg pattern was the reverse of this modern finding: heights converged to the levels of the developed regions, but did not converge among the more developed regions themselves. Further work needs to explore the reasons for the spatial convergence in heights, even if limited. Was it government redistribution in favor of the periphery, the diffusion of provision of public health services, railroad construction or income and price trends? To a certain extent, the phenomenon might also have biological aspects, insofar as physical growth is non-linear, so that taller populations are expected to increase their heights at a slower rate.

All in all, the evidence on heights is similar to the evidence on GDP growth in the Habsburg Monarchy insofar as they both point to some positive elements but are by no means uniformly favorable. Evidently, the situation in the Monarchy was by no means hopeless before its demise, but it was also not among the high achievers of the time as the Scandinavian countries which also began the period under considerations as peripheral economies.

References

Austria. *Militär-Statistisches Jahrbuch für das Jahr...*, (Annually, 1870-1910). Vienna, k. k. Hof- und Staatsdruckerei, 1873-1913.

Baumol, William J. 1986. “Productivity Growth, Convergence and Welfare: What the Long-Run Data Show.” *American Economic Review* 76: 1072-85.

Barro, Robert J. and Sala-i-Martin, Xavier, 1995, *Economic Growth*. New York: McCraw Hill.

Brabec, M., 2005: Analysis of periodic fluctuations of the height of Swedish soldiers in 18th and 19th centuries. *Economics and Human Biology*. Volume 3, Issue 1, Pages 1-16.

Cox, D.R. and Oakes, D., 1984, *Analysis of Survival Data*. London. Chapman and Hall.

Cressie, N., 2001, *Statistics for Spatial Data*. New York. Willey.

Davidian, M. and Giltinan, D.M. 1995, *Nonlinear Models for Repeated Measurement Data*. London. Chapman and Hall.

Galor, Oded. 1996. "Convergence? Inferences from Theoretical Models." *Economic Journal* 106, (437): 1056-1069.

Good, David F., 1984. *The Economic Rise of the Habsburg Empire, 1750-1914*, Berkeley: University of California Press.

Hastie, T., Tibshirani, R. and Friedman, J. 2001, *The Elements of Statistical Learning*. New York. Springer.

Komlos, John *The Habsburg Monarchy as a Customs Union: Economic Development in Austria-Hungary in the Nineteenth Century*, Princeton University Press: 1983.

Johnston, J. 1984, *Econometric Methods*. New York. McGraw-Hill.

Longford, N.T. 1995, *Random Coefficient Models*. Oxford. Oxford.

Meeker, W.Q. and Duke, S.D., 1981, CENSOR – A User-Oriented Computer Program for Life Data Analysis. *The American Statistician*, 35, 112.

Pinheiro, J.C. and Bates, D.M. 2000, *Mixed-Effects models in S and S-PLUS*. New York. Springer.

Schulze, Max-Stephan, 2000, Patterns of Growth and Stagnation n the late Nineteenth Century Habsburg Economy. *European Review of Economic History*, 4, 311-340.

Endnotes

1 The two numbers separated by a comma refer to the range of the bins.

2 See e.g. Brabec (2005).

[3] In this paper, we will use the following coding for the 15 districts: 1=Vienna, 2=Prague, 3=Gratz, 4= Innsbruck, 5=Zara (Dalmatia), 6=Lemberg (Galicia), 7=Budapest, 8=Hermannstadt (Transylvania), 9=Zagreb (Croatia), 10=Krakau (Galicia), 11=Josefstadt (Bohemia), 12=Pressburg (Slovakia), 13=Kaschau (Slovakia), 14=Temesvar (Transylvania), 15=Przemysl (Galicia).

[4] SEM = Standard Error of the Mean. Due to censoring, its computation is slightly more involved and values typically somewhat larger than those obtained from the standard formula: $SEM = \frac{1}{\sqrt{n}}$.

[5] See e.g. Longford (1995), Davidian and Giltinan (1995) for general exposition.

[6] See Pinheiro and Bates (2000) for more information on REML estimates and their computation. 0.203 is residual standard deviation, lumping together all deviations from the model. 1.525 is the sd describing variability of initial conditions across districts, 0.036 is the sd describing variability of district-specific slopes (or 1890-1910 average secular velocities).

[7] See Hastie, Tibshirani and Friedman (2001) for introduction to local regression.

[8] See Cressie (1991) for a comprehensive exposure to spatial statistics.

[9] The likelihood can take different values of (vector) parameter as argument and here, the value equal to the MLE was taken.

[10] If different lags were orthogonal (which they are not), the result would be the same as that from the model having all 21 lag values for a particular indicator. Note that the model with free coefficients on all 21 lags is not only unstable here, it is even not identifiable. The approach here is to introduce the problem in a simple way (one might think of using a plot like this if a lag with highest coefficient would be sought for – as in the situation when one wants to „optimize“ but only one lag is allowed to enter the regression). Something else needs to be done if one wants to proceed further. And that is done via Almon lag structure, below.

[11] Note that we did not attempt simultaneous optimization in 3 dimensions across all combinations of all three lags.

[12] The model would be as good if we dropped industrial and agricultural output. The likelihood ratio test between (10) and its restriction having only lag 2 population as explanatory variable is not significant (p=0.811). The regression coefficient on the lag 2 population is slightly lower in absolute value in the restricted model as compared to (10).

[13] After World War II life expectancy has been converging around the globe even though income has not been (Kenny 2005).