

Abstract. A long series of attempts have been made to identify factors influencing the number of males per 100 females at birth, also called the secondary sex ratio (SR). Earlier studies have noted regional variations in the secondary SR, and racial differences seem to exist. Statistical analyses have shown that comparisons between SRs demand large data sets because for moderate data the random fluctuations are marked, and hence, reliable results presuppose national birth data. Variations in the SR that have been reliably identified in family data have mostly been slight and without notable influence on national birth registers. In this study, we analyse the regional variations in the SR in Sweden (1749-1869). We build spatial models for the regional variations in the SR and, in addition, analyse the association between the SR and the crude birth rate (CBR) and the total fertility rate (TFR). The strong random fluctuations in the SR yielded rather poor goodness of fit. Especially, the proposed models failed to satisfactorily identify counties with observed extreme SRs.

Key words: live births; counties; geographical co-ordinates; spatial models; stepwise regression; multicollinearity; crude birth rates; total fertility rates

Introduction

In a long series of papers, attempts have been made to identify factors influencing the number of males per 100 females at birth, also called the secondary sex ratio (SR). Hawley [1] stated that where prenatal losses are low, as in the high standard of living in Western countries, the SRs are usually around 105 to 106. On the other hand, in areas with a lower standard of living, where the frequencies of prenatal losses are relatively high, SRs are around 102. Visaria [2] stressed that racial differences appear to exist in the SR. Furthermore, he noted that the observed SRs are strongly influenced by random errors, and he constructed confidence intervals (CIs) for the SR. Fellman and Eriksson ([3], [4]) presented an alternative CI

$$\left(SR - \frac{k}{1-p} \sqrt{\frac{p}{(1-p)N}}, SR + \frac{k}{1-p} \sqrt{\frac{p}{(1-p)N}} \right), \quad (1)$$

29 where N is the sample size, p is the proportion of males, $SR = \frac{p}{(1-p)}$ and k defines the
 30 confidence level (e.g. $k = 1.96$ yields 95% CIs). In empirical studies, the SR is commonly
 31 multiplied by 100, yielding the number of males per 100 females. The obtained CIs are
 32 asymptotically equivalent and both are applicable for large N . For moderate data sets, the CIs
 33 are broad, and consequently, it is difficult to identify statistically significant differences ([2];
 34 [3]). Fellman and Eriksson [3] analysed in detail the SR among singletons, twins and higher
 35 multiple births in Sweden (1869-1967). Later, Fellman and Eriksson [4] analysed the
 36 temporal variations in the SR in Sweden, Denmark, Finland, Norway and Iceland.
 37 Variations in the SR that have been reliably identified in family data have in general been
 38 slight and without notable influence on national birth registers (for references, see [5], [3]).

39 **Material**

40 In the 19th century, Berg [6] published SR data for live births in the counties of Sweden,
 41 1749-1869 (see Figure 1 and Table 1). The available periods varied slightly between the
 42 counties. The regional data provide the basis for our study. We did not consider subperiod
 43 data for the counties because SR data for small data sets show large random fluctuations ([2],
 44 [3]). Berg defined the SRs as males per 1000 females, but we have transformed his data to the
 45 traditional definition, number of males per 100 females.

46 Several studies have shown marked temporal variations. A review of studies of temporal
 47 variations in the SR is given in [3]. For the period 1749-1869, such temporal trends seem to
 48 be minute. In Figure 2, we present the temporal variations in the SR in Sweden (1751-1870).
 49 The increasing trend seems to be slight. The temporal trend in the SR for Sweden is discussed
 50 in [4].

51 **Figure 1**

52 **Figure 2**

53 **Methods**

54 The first attempt presented by Fellman and Eriksson [8] was to trace any regional
 55 heterogeneity in the SR. This was performed so that the numbers of males and females in the
 56 counties were estimated by the total numbers of births and the published regional SRs, both

given by Berg. Significant differences in the sex proportions were found, and their next step was to build regional models.

Following the ideas in Fellman and Eriksson [9], we applied spatial regression models to the regional SRs. The location of the counties was defined as the geographical co-ordinates of the corresponding residences (provincial capitals). The residences are not centrally located, but we assumed that they are sufficiently central with respect to the population density, and their co-ordinates are given in Table 1. The county codes introduced by Statistics Sweden are presented in Figure 1 and used in Table 1. The co-ordinates for Sweden are eastern longitude and northern latitude. Table 1 also displays for every county the observation periods, the number of live births and the observed and estimated SRs. We also introduced fertility regression models based on the crude birth ratio (CBR) and the total fertility ratio (TFR).

Table 1

Spatial models. We analysed the spatial variation in the SR by weighted regression models. The variance of the SR is approximately proportional to n^{-1} , and therefore, we used the number of births in the counties as weights. The regressand was the observed regional SR, the presumptive regressors were the longitude (meridian) M and the latitude L and the transformed variables were L^2 , M^2 and LM . The regressors M and L were defined as deviations from the co-ordinates of the unweighted centre (59.18° N and 15.87° E) of the cluster of residences, and consequently, the intercepts obtained in the models are the estimates of the SR in this centre.

Figure 1

The geographical pattern of Sweden can be seen in Figure 1. The elongated shape of Sweden indicates that attention must be paid to the multicollinearity between the regressors. Fellman and Eriksson [9] studied spatial models for the twinning rate in Sweden and gave a thorough analysis of the multicollinearity in the regression models with the geographical co-ordinates as regressors. They considered different measures of multicollinearity proposed in the literature. These were all functions of the eigen values of the correlation matrices. In this

study, we restrict ourselves to one measure, that is $m_4 = \sum_i \frac{1}{\lambda_i}$. This measure includes all

eigen values and for uncorrelated regressors it equals the dimension of the correlation matrix and with increasing multicollinearity it increases towards infinity. If all regressors are included, the multicollinearity is rather strong, but with a smaller set of regressors the

multicollinearity is markedly reduced. For the total model $m_4 = 38.09$. The optimal spatial regression model presented below has only the regressors L^2 and LM , and in this model the multicollinearity measure is reduced to $m_4 = 13.74$. This model is called “the spatial model”.

The measure ($m_4 = 32.56$) for the total model presented in [9] differs from the value obtained here since in the earlier paper the county of Örebro was missing. In addition, the number of births was obtained for different periods, and consequently, the weights were slightly different.

Fertility models. Jalavisto [10] pointed out that in Finland the sex ratio has risen continuously and steadily from 1751 to 1948, and she showed that the lower rate of male births in the past could not be attributed to higher mean parity.

Vartiainen et al. [11] evaluated the SR among live births in Finland, 1751-1997. They observed an increasing trend from 1751 to 1920; this was followed by a decrease and interrupted by peaks during and after World War I and II. None of the family parameters (paternal age, maternal age, age difference of parents, birth order) could explain the time trends. Furthermore, a brief rise occurred in the early 1970s. The last peak is more difficult to explain. The authors noted that the peak was associated with reduced birth rates and an increased proportion of first-born children.

Following [10] and [11], we tried to connect the variations in the SR with two measures of fertility, CBR and TFR. CBR is the number of childbirths per 1000 people per year. Hofsten and Lundström [12] presented in their Table 6.1 the CBRs for the counties in Sweden for the decades between 1751 and 1970. They stressed that the CBR is a poor measure of fertility and can only be used for rough comparisons or for comparisons when it is known that no other factors – notably, the age composition of the population, the marriage pattern or the time-schedule for the birth of children – will interfere with the comparison. However, they stated that differences between the counties were minor before the transition started from high to low fertility, and they assumed that this change occurred during the later part of the 19th century. We therefore assume that CBR can be used in this study of regional comparisons for the period 1749-1869.

The TFR of a population is the average number of children born to a woman over her lifetime if (1) she were to experience the exact current age-specific rates through her lifetime, and (2) she were to survive from birth to the end of her reproductive life. TFR is obtained by summing the single-year age-specific birth rates at a given time. Hofsten and Lundström [12]

have in Tables 6.2 – 6.16 presented TFR values per 1000 women for all decades starting from around 1860 to 1970.

In this study, we define the variable CBR as the mean value of the decennial CBR data for the period 1751-1870, and the variable TFR is the Hofsten-Lundström data for 1860. We calculate for the regional values the means $\overline{CBR}$ and $\overline{TFR}$. In our regression analyses, we use as regressors the variables $C = CBR - \overline{CBR}$ and $T = TFR - \overline{TFR}$. This model is named “the fertility model”.

Both C and T are in some sense measures of fertility, and the correlation between them is 0.549. The correlation results in the multicollinearity measure $m_4 = 3.72$.

Results

In general, for moderate data sets, the SR is influenced by large random fluctuations ([2], [3], [4]). This can be seen in Figure 3, where we present the regional SRs with 95% CIs. Note the broad CIs for the counties of Västerbotten (AC), Gävleborg (X), Västernorrland (Y), Norrbotten (BD), Gotland (I) and Jämtland (Z). For these, the number of live births is less than 200 000.

Figure 3

Spatial models. The test of the regional variations in SR yielded $\chi^2 = 55.23$ with 24 degrees of freedom, indicating regional heterogeneity. Our first attempt to build a spatial model started with the maximal model

$$SR = \alpha + \beta_M M + \beta_L L + \beta_{M^2} M^2 + \beta_{LM} LM + \beta_{L^2} L^2 + \varepsilon. \quad (2)$$

The estimated model was

$$SR = 104.58 - 0.085 M + 0.088 L - 0.0118 M^2 - 0.069 LM + 0.080 L^2. \quad (3)$$

No parameter estimate was significant, the adjusted coefficient of determination was low, $\overline{R}^2 = 0.047$, and consequently, the goodness of fit was poor.

We then improved the analyses by a stepwise regression analysis, yielding the optimal spatial model, that is the model with a maximal $\overline{R}^2$ value. The model obtained contained L^2 and LM and was

$$SR = 104.56 - 0.067 LM + 0.068 L^2. \quad (4)$$

The goodness of fit was improved, but the adjusted coefficient of determination $\bar{R}^2 = 0.103$ still indicated a rather poor fit. The estimated parameter $\hat{\beta}_{L^2}$, with $SD_{L^2} = 0.031$, was statistically significant, and $\hat{\beta}_{LM}$ ($SD_{LM} = 0.034$) was almost significant. The intercept $SR = 104.56$ is an estimate of the SR in a hypothetical county with latitude = 59.18 and longitude = 15.87. The county of Örebro with the residence Örebro comes closest to this cluster centre (cf. Table 1).

Fertility models As an alternative model, we consider a regression model with the regressors C and T , that is

$$SR = \alpha + \beta_C C + \beta_T T + \varepsilon. \quad (5)$$

The estimation yields the regression model

$$SR = 104.68 - 0.1445 C + 0.00086 T. \quad (6)$$

Compared with the optimal spatial model, the goodness of fit is better because the adjusted coefficient of determination $\bar{R}^2 = 0.373$ and the parameter estimates are both statistically significant, with $SD_C = 0.04867$ and $SD_T = 0.00024$, respectively. Note that the regression coefficient $\beta_C = -0.1445$ is negative. This finding supports to some extent Jalavisto's statement that the increased SR in her data in the 1970s is associated with a low birth rate. The intercept $SR = 104.68$ is an estimate of the SR in a hypothetical county with $CBR = 32.599$ and $TFR = 4574.64$. The county of Älvsborg comes closest to these values (cf. Table 1).

The goodness of fit of both models was poor. Together with the observed SRs, the estimated SRs for the spatial model and the fertility model are given in Figure 4. Figure 4 shows that no model can satisfactorily explain the broad heterogeneity in the regional SR. The most marked discrepancies between the observed and estimated SRs are seen among counties with low SRs for the spatial model and among counties with high SRs for the fertility model (Figure 5). The difference in the discrepancies between the two models raises the question that would a combination of the two models improve the fit. However, no improvement was obtained when spatial variables were added to the fertility model.

175 **Figure 4**

176 **Figure 5**

177 **Discussion**

178 The slight increasing tendency in the SR for Sweden (1751–1870) observed in Figure 2 may
 179 be related to an increased standard of living ([1], [4]). In Figure 2, one can also see a minor
 180 trough during the first half of the 19th century. During this period the standard of living was
 181 temporarily low in Sweden. Sandberg and Steckel [7] used measurements of Swedish soldiers
 182 recorded in muster rolls as an index of living standards and noted a reduction in mean heights.
 183 Their results combined with other evidence point to a mild subsistence crisis, especially in
 184 western Sweden, during this period. Eriksson et al. [13] discussed *inter alia* the connection
 185 between this crisis and the twinning rate in Sweden and concluded that low values in the
 186 twinning rate are not consistently associated with periods of epidemics, famine or other
 187 similar nutritional stress.

188 When one considers the counties with high and low SRs, no obvious association with the
 189 standard of living emerges. Among counties with high SRs are the northern counties of
 190 Jämtland (Z), Norrbotten (BD) and Västernorrland (Y), which were at that time relatively less
 191 developed parts of Sweden. On the other hand, among the counties with low SRs are the city
 192 of Stockholm (A), the county of Göteborg and Bohus (O) with the second largest city of
 193 Gothenburg, and the county of Uppsala (C), which contains the oldest university in Sweden.
 194 These counties seem to be among the most developed parts of the country. However,
 195 Sandberg and Steckel [7] stressed that mid-19th century Stockholm emerges as the markedly
 196 unhealthy city that historical chronicles and demographic data have always portrayed. Among
 197 counties with low SR is the northern county of Västerbotten (AC). Note that Västerbotten is a
 198 neighbouring county to Jämtland and Norrbotten.

199 No geographical pattern can be detected. Counties with low SRs comprise western (Göteborg
 200 and Bohus), eastern (city of Stockholm, Uppsala, Västmanland), southern (Blekinge) and
 201 northern (Västerbotten) counties. Included among counties with high SRs are western
 202 (Värmland), eastern (Gotland, Kalmar) and northern (Västernorrland, Jämtland, Norrbotten)
 203 counties. This lack of geographical pattern supports the poor goodness of fit of the spatial
 204 model. Figure 1 gives summary information about the geographical pattern of Sweden.

205 Concerning the fertility model, we have already connected the negative $\hat{\beta}_C$ with Jalavisto's
 206 result. The variables CBR and TFR are positively correlated. Despite this, both variables

must be included in the model, one with a negative coefficient ($\hat{\beta}_C$) and the other with a positive coefficient ($\hat{\beta}_T$). Among the counties with low CBRs, Jämtland (Z) and Gotland (I) have high SRs, but Uppsala (C) has a low SR. Among the counties with high CBRs, city of Stockholm (A), Blekinge (K) and Göteborg and Bohus (O) have low SRs, but Norrbotten (BD) has a high SR. Among counties with high TFRs, Norrbotten (BD) and Västernorrland (Y) have high SRs, but the neighbouring county Västerbotten (AC) has a low SR. Among counties with low TFRs, city of Stockholm (A) and the county of Uppsala (C) have low SRs, but Gotland (I) has a high SR. These tendencies are supported by the signs of the parameters, but some exceptions exist.

Hofsten and Lundström [12] stressed that as early as about 1860 the city of Stockholm and the county of Gotland displayed a fertility considerably lower than that for the country overall. The difference being most marked in the higher age groups seems to indicate an early influence of birth control.

Conclusions

The slight increasing tendency in the SR for Sweden (1751-1870) observed in Figure 2 may be related to an increased standard of living ([1], [4]). In Figure 2, one can also see a minor trough during the first half of the 19th century. During this period the standard of living was temporarily low in Sweden. Sandberg and Steckel [7] used measurements of Swedish soldiers recorded in muster rolls as an index of living standards and noted a reduction in mean heights. Their results combined with other evidence point to a mild subsistence crisis, especially in western Sweden, during this period. Eriksson et al. [13] discussed *inter alia* the connection between this crisis and the twinning rate in Sweden and concluded that low values in the twinning rate are not consistently associated with periods of epidemics, famine or other similar nutritional stress.

References

- [1] Hawley, A. H. (1959). Population composition. In *The Study of Population: An Inventory and Appraisal*. Ed. P. M. Hauser and O. Dudley Duncan. Chicago, University of Chicago Press, 361-382.
- [2] Visaria, P. M. (1967). Sex ratio at birth in territories with a relatively complete registration. *Eugen. Quart.* 14:132-142.

- [3] Fellman, J. & Eriksson, A. W. (2010). Secondary sex ratio in multiple births. *Twin Res. Hum. Genet.* 13(1):101-108
- [4] Fellman, J. & Eriksson, A. W. (2011). Temporal Trends in the Secondary Sex Ratio in Nordic Countries. *Biodemography and Social Biology*, 57 (2):143-154, DOI: [10.1080/19485565.2011.614193](https://doi.org/10.1080/19485565.2011.614193).
- [5] Fellman, J., Eriksson, A. W. & Forsius H. (2002). Sex ratio and proportion of affected sons in sibships with X-chromosomal recessive traits: Maximum likelihood estimation in truncated multinomial distributions. *Hum. Hered.* 53:173-180.
- [6] Berg, F. T. (1871). Proportionen mellan könen bland de födde och inom den stående befolkningen med hänsyn till Sverige och dess provinciella olikheter. (Sex ratio at birth and in the population with respect to Sweden and its regions). *Kungliga Svenska Vetenskaps-Akademiens Handlingar* 10(6):1-40 and 3 figurer.
- [7] Sandberg, L. G. & Steckel, R. H. (1987). Heights and economic history: the Swedish case. *Annals of Human Biology*, 14:101-110.
- [8] Fellman, J. & Eriksson, A. W. (2014). Effect of fertility on secondary sex ratio and twinning rate in Sweden, 1749-1870. *Twin Research and Human Genetics*, 18(1):92-99. doi:[10.1017/thg.2014.71](https://doi.org/10.1017/thg.2014.71)
- [9] Fellman, J. & Eriksson, A. W. (2009). Spatial variation in the twinning rate in Sweden, 1751-1850. *Twin Res. Hum. Genet.* 12(6):583-590
- [10] Jalavisto, E. (1952). Sex ratio at birth and its dependence upon birth order and parental age. *Ann. Chir. Gynaec. Fenn.*, 41:129-137.
- [11] Vartiainen, T. Kartovaara, L. & Tuomisto, J. (1999). Environmental chemicals and changes in sex ratios: Analysis over 250 years in Finland. *Environmental Health Perspectives*. 107:813-815.
- [12] Hofsten, E. & Lundström, H. (1976). *Swedish Population History. Main trends from 1750 to 1970*. Urval 8. National Central Bureau of Statistics. Stockholm, 186 pp.
- [13] Eriksson, A. W., Bressers, W. M. A., Kostense, P. J., Pitkänen, K. J., Mielke, J. H. Jorde, L. B., Tas, R. F. J. & Fellman, J. O. (1988). Twinning rate in Scandinavia, Germany and the Netherlands during years of privation. *Acta Genet. Med. Gemellol.* 37:277-297.
- Table 1.** County codes, observation periods, residences, geographical co-ordinates of the residences, number of births and observed and estimated secondary sex ratios (SRs) among live births in the counties of Sweden. The row Total includes the geographical co-ordinates of

the centre of the cluster of residences, the total number of births, the observed SR of Sweden overall and the intercepts in the spatial and fertility models. For details, see the text.

273

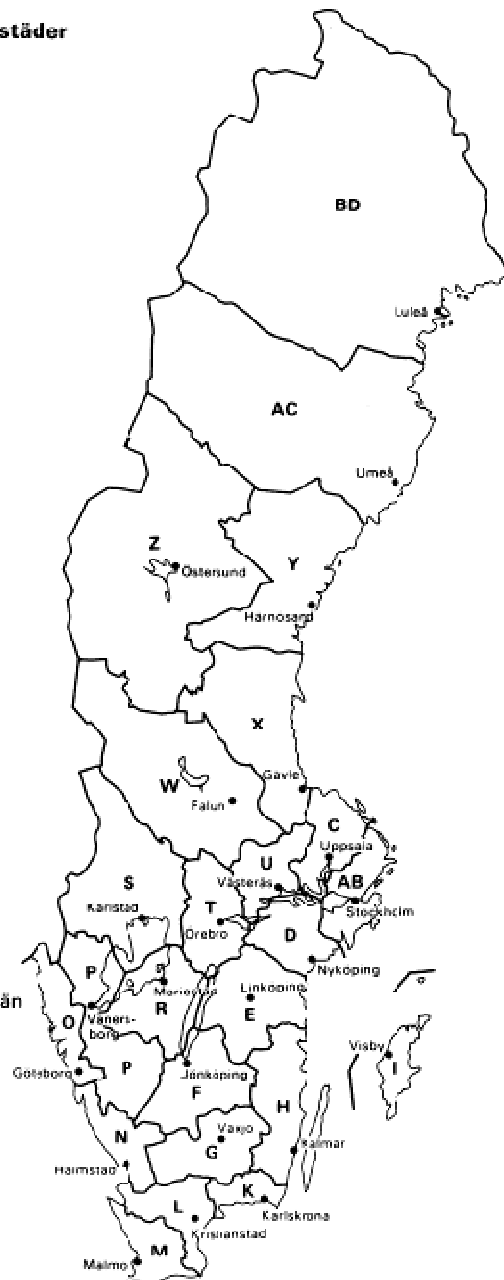
Code ^{a)}	Period ^{b)}	Residence	Lat	Long	n	SR obs	SR spat	SR fert
A	1749-1869	Stockholm	59.32	18.07	336854	103.4	104.54	103.69
B	1749-1869	Stockholm	59.32	18.07	324901	104.6	104.54	104.46
C	1749-1869	Uppsala	59.90	17.80	246344	104.2	104.51	104.49
D	1749-1869	Nyköping	58.76	17.01	319940	104.6	104.61	104.81
E	1749-1869	Linköping	58.42	15.64	581692	104.5	104.59	104.65
F	1749-1869	Jönköping	57.78	14.18	424184	104.9	104.54	104.97
G	1749-1869	Växjö	56.86	14.82	356405	104.7	104.77	104.79
H	1749-1869	Kalmar	56.80	16.00	516021	105.8	104.97	104.72
I	1749-1869	Visby	57.63	18.30	119541	105.3	104.98	104.50
K	1749-1869	Karlskrona	56.16	15.58	283511	103.7	105.13	104.46
L	1749-1869	Kristianstad	56.02	14.13	455200	104.8	104.88	104.70
M	1749-1869	Malmö	55.61	13.06	637249	104.6	104.76	104.45
N	1749-1869	Halmstad	56.67	12.86	271859	104.5	104.49	104.82
O	1749-1869	Göteborg	58.35	11.93	482251	103.9	104.39	103,81
P	1749-1869	Vänersborg	58.37	12.32	597113	104.8	104.42	104.70
R	1749-1869	Mariestad	58.71	13.82	522657	104.8	104.51	104.92
S	1749-1869	Karlstad	59.38	13.50	552016	105.1	104.60	104.85
T	1749-1869	Örebro	59.27	15.22	367499	104.5	104.57	105.12
U	1749-1869	Västerås	59.67	16.55	272943	104.3	104.56	104.61
W	1749-1869	Falun	60.61	15.64	395484	104.7	104.72	105.04
X	1810-1869	Gävle	60.68	17.16	188398	104.8	104.59	104,77
Y	1810-1869	Härnösand	62.63	17.94	175813	105.1	104.90	104.93
Z	1810-1869	Östersund	63.18	14.65	77473	106.4	105.98	105,34
AC	1749-1869	Umeå	63.83	20.27	169733	104.4	104.67	104.68
BD	1749-1869	Luleå	65.59	22.17	153877	105.1	104.67	104.79
Total	1749-1869		59.18	15.87	8828958	104.66	104.56	104.68

274 a The codes are explained in Figure 1.

275 b) For Stockholm city (A) and the county of Gotland (I), the data are available for the
 276 whole period, but for the rest of the counties there are missing data for the period 1774-1794.

Sveriges län och residensstäder
*Counties and provincial
 capitals of Sweden*

- AB = 01 Stockholms län
 C = 03 Uppsala län
 D = 04 Södermanlands län
 E = 05 Östergötlands län
 F = 06 Jönköpings län
 G = 07 Kronobergs län
 H = 08 Kalmar län
 I = 09 Gotlands län
 K = 10 Blekinge län
 L = 11 Kristianstads län
 M = 12 Malmöhus län
 N = 13 Hallands län
 O = 14 Göteborgs och Bohus län
 P = 15 Älvsborgs län
 R = 16 Skaraborgs län
 S = 17 Värmlands län
 T = 18 Örebro län
 U = 19 Västmanlands län
 W = 20 Kopparbergs län
 X = 21 Gävleborgs län
 Y = 22 Västernorrlands län
 Z = 23 Jämtlands län
 AC = 24 Västerbottens län
 BD = 25 Norrbottens län



8 Geografiska uppgifter *Geographical data*

Figure 1. Map of Sweden including the counties (län) and their residences (provincial capitals) and the letter codes according to Statistics Sweden.

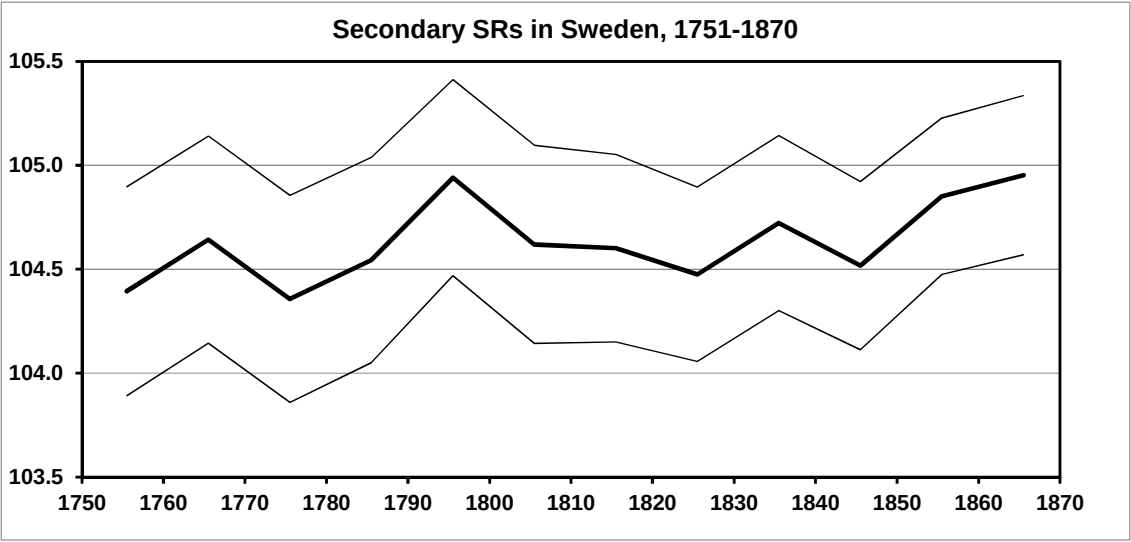
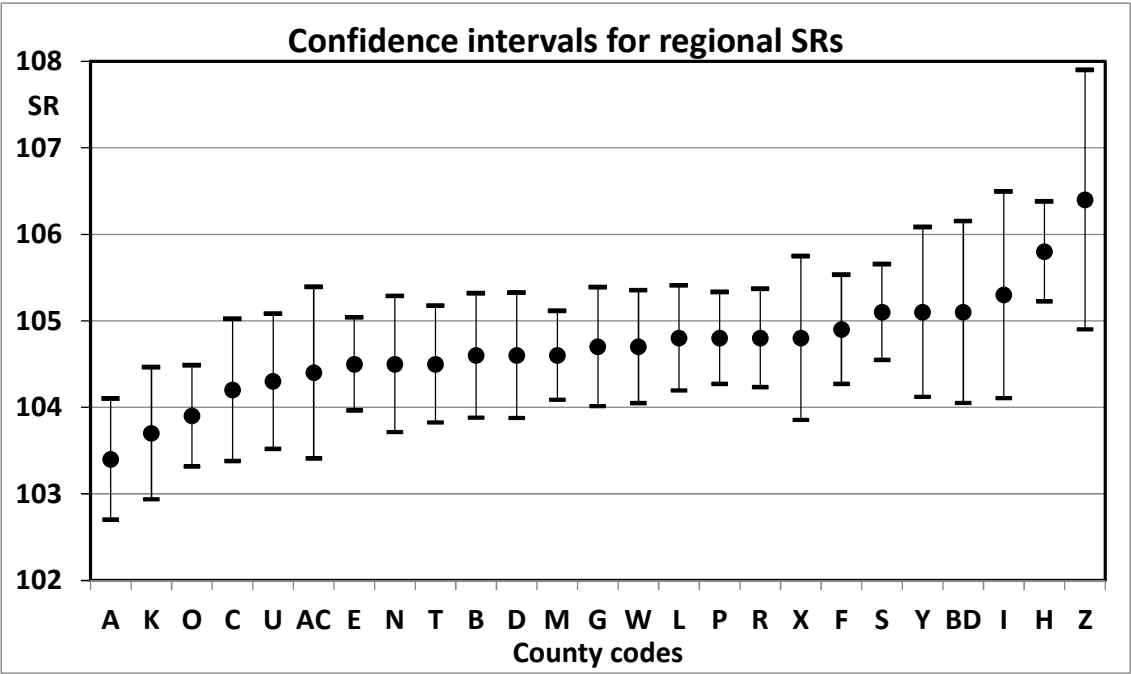


Figure 2. Secondary sex ratios (SRs) among live births in Sweden, 1751-1870. Included in the figure is a 95% confidence band. A slight increasing trend can be noted, which is probably associated with the increasing standard of living in Sweden (cf. Fellman & Eriksson, [4]). A minor trough can be seen between 1800 and 1860. Sandberg and Steckel [7] noted a reduction in the mean heights among Swedish conscripts. Their results combined with other evidence point to a mild subsistence crisis in western Sweden during this period.



290 **Figure 3.** Observed secondary sex ratios (SRs) and confidence intervals for different counties.
291 The counties are ordered according to increasing SR, and the county codes are given in
292 Figure 1.
293

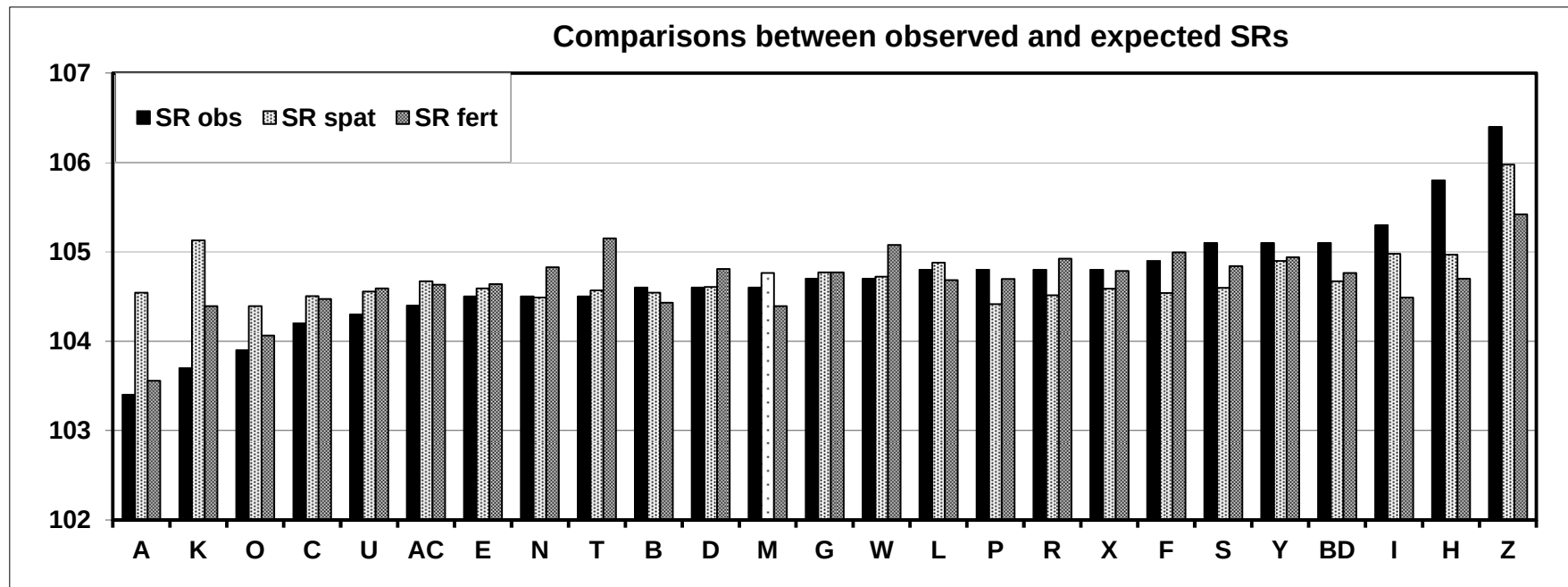
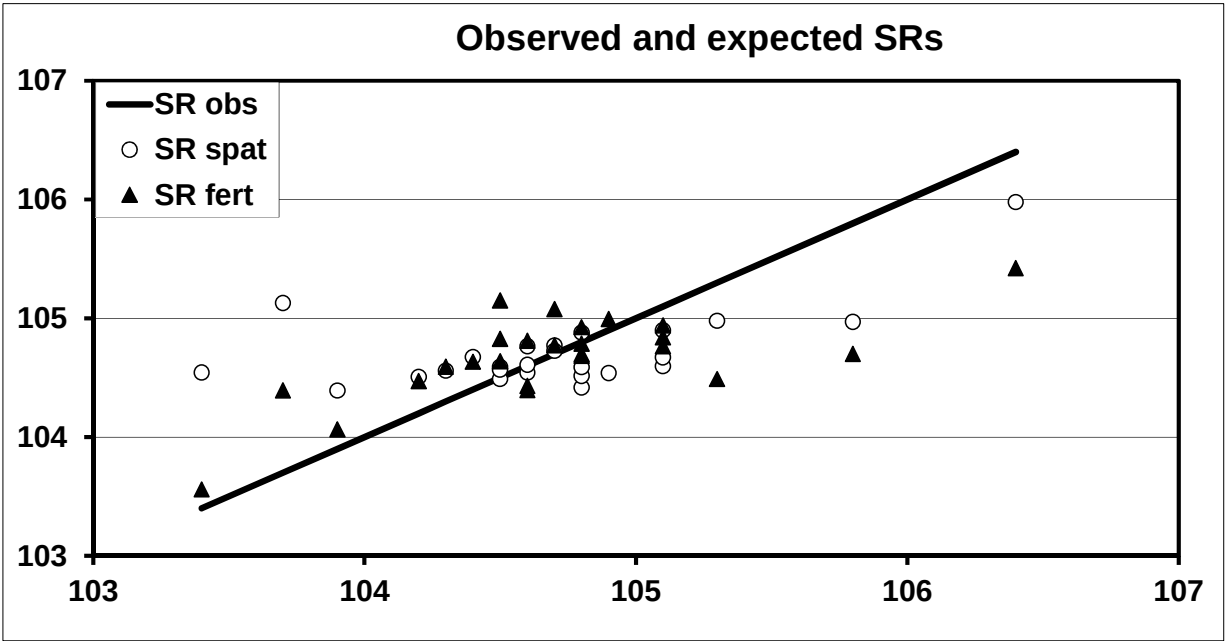


Figure 4. Observed and estimated secondary sex ratios (SRs). The applied models are the spatial (SR spat) and fertility (SR fert) models. Note that the extreme observed secondary sex ratios (SR obs) are poorly estimated. The counties are ordered according to increasing observed SR values, and the county codes are given in Figure 1.

300



301

302

303 **Figure 5.** Comparisons between observed secondary sex ratios SRobs and estimated SRs for
304 the spatial (○) and fertility (▲) models for the counties in Sweden. Note the discrepancies
305 for counties with observed extreme SR.