Online Appendix S2: Technical Appendix

This appendix expands on two of the potential problems of bias that we face in using probate records: variation in death rates across sectors and wealth bias.

(a) Variation in Adult Death Rates

One source of potential bias in the probate record stems from the fact that if the adult death rate differs across sectors, sectoral shares amongst the dead and hence in the probate record will only imperfectly mirror sectoral shares in society at large. As sectors with a high (low) death rate will be over- (under-) represented, the size of the bias increases with the ratio between the sector-specific death rate and that in society at large. Formally:

$$s_s = \left(\frac{\omega_s}{\omega}\right)^{-1} \vartheta_s \,, \qquad \text{(A1)}$$

where s_s is the sectoral share in sector s (agriculture, industry, and services, respectively), ω_s is the adult death rate in sector s, un-subscripted ω is the death rate in society at large, and θ_s is the sectoral share in sector s in the probate record.

How big was this bias in the context of early modern England? Here we have to consider two effects: income effects are expected to cause a positive bias in the agriculture share; the urban mortality penalty should have an opposite effect. As our agricultural share is comparatively low, we are particularly concerned about the latter effect and indeed Wrigley et al. (1997, 202–203) argue that in early modern England the urban/rural divide mattered more than income for mortality. A comparison between their estimates and those reported by Woods (2003, p. 36) suggests that the ratio between life expectancy in the countryside and in the city is a good guide to ratios between mortality rates: for Woods, life expectancy in the early modern English countryside was about 1.5 times that in the city; for Wrigley et al., levels of mortality in the city may have been 60 percent higher than in the countryside. Much of the difference was due to infant mortality and Woods' (2003, p. 36) figures also imply that life-expectancy at 15—which is the relevant one for the work-force—in early modern London was about 90 percent of that of England. If anything we expect the London figure to provide an upper bound of the urban penalty, given that this increased with population density. The figures thus suggest that for adults (people aged 15 or more) c. 1.10 (as 1/0.9=1.11) is a reasonable estimate of mortality rates in the city relative to England.

This is confirmed also by available data on age-specific mortality probabilities. Landers (1993, p. 172) reports figures for London in 1730–1749, which can be compared with those reported for England at the same time by Wrigley et al. (1997, pp. 262, 290), interpolating as they do with Brass' (1971) method for the intervals between 15 and 25. Comparison with adult life expectancy in other decades (Wrigley et al. 1997, p. 290) suggests that there is nothing unusual about 1730–1749 in terms of mortality. Within each age-interval the crude death rate is the mortality probability divided by the length of the interval. The overall adult crude death rate in London and England is the weighted average of the crude death rates of the intervals from age 15 onwards, where the weights are given by the proportion of the adult population covered by each age interval. The latter are drawn from Wrigley and Schofield's (1989, p. 218) estimates of the age structure of England in 1696 (assuming for simplicity that nobody is older than 75). The resulting adult crude death rates in London and England are 25.60 per thousand and 21.59, yielding a ratio of 1.18.

The crude death rate in England is equal to the weighted average between the death rate in the cities and that in the countryside, with the weights defined by the urbanization rate. It follows that:

$$\omega_r = \frac{\omega - \omega_u U}{(1-U)} \tag{A2}$$

where ω_r and ω_u are the adult death rates in the countryside and in the city, respectively, U is the urbanization rate. Given that in our period on average the urbanization rate was about 14 percent, equation (A1) and the estimated adult crude death rates imply an adult crude death rate in the countryside of 20.93 per thousand, which yields a rate with the national one of 97 percent. The following graph compares our agricultural share (balanced sample) with that adjusted for different death rates in the countryside and the city using this value and equation (A1):

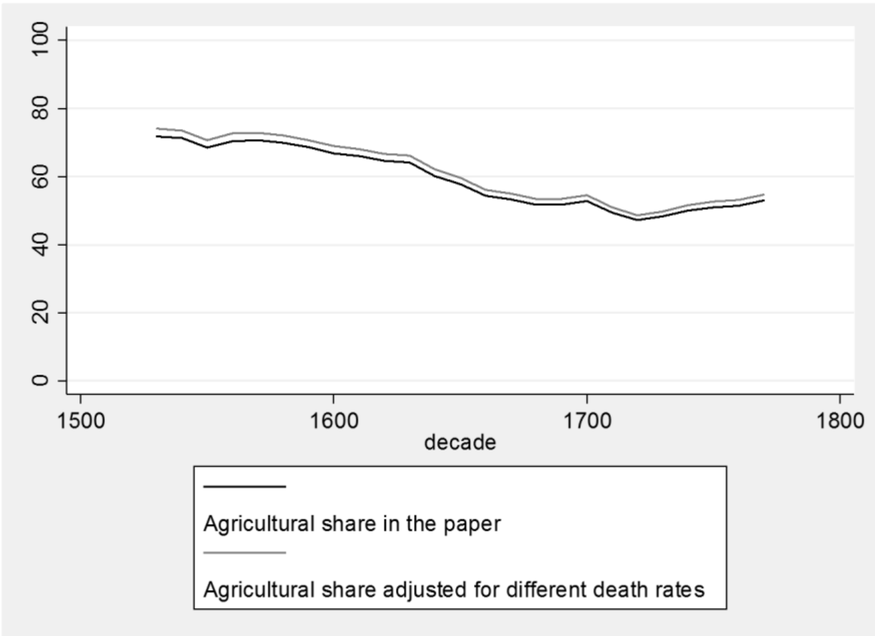

FIGURE S2.1
AGRICULTURAL SHARE IN THE PAPER AND ADJUSTING
FOR DIFFERENT ADULT DEATH RATES
IN THE COUNTRYSIDE AND THE CITY

Sources: See Appendix 1 and 2 and the text.

The difference is very modest (less than 2 percentage points on average) and the two plots tell the same story. It is therefore safe to neglect the urban mortality penalty for the purpose of the analysis, particularly as this is an upper bound of the adjustment needed.

(b) *Wealth Bias in the Probate Record*

The key source of bias in the probate record is that only people with wealth and/or capital to bequeath are likely to appear in it. The size of this bias increases with the ratio between the fraction of deaths leaving a will in each sector and that in society at large. Formally:

$$s_s - \vartheta_s = s_s \left[1 - \frac{\delta_s}{\sum_s \delta_s s_s}\right], \tag{A3}$$

where s_s is the sectoral share in sector s, θ_s is the sectoral share in sector s in the probate record, and δ_s is the share of deceased leaving a will in sector s. Only the society-wide fraction of deaths leaving a will, $\sum_s \delta_s s_s$, is known. However, the sector-specific fractions of deaths with a probate record can be computed on the basis of how sectoral shares in the probate record vary with the share of deaths leaving a will. In fact, the relationship between the two variables can be estimated by running the following regression:

$$\theta_{sit} = exp(\alpha_{si} + \beta_{1si}t + \beta_{2si}t^2 + \beta_{3s}\delta_{it})/ \\ [1 + exp(\alpha_{si} + \beta_{1si}t + \beta_{2si}t^2 + \beta_{3s}\delta_{it})] + u_{sit}, \tag{A4}$$

where θ_{sit} is the sectoral share from the probate record in sector s, county i and time t, α_{si} is a county/sector constant, β_{1si} and β_{2si} are the coefficients of the county/sector quadratic trends, β_{3s} is the main coefficient of interest, and δ_{it} is the share of deaths covered by the probate record. Neglecting for simplicity the error, if δ_{it} is 100 percent then θ_{sit} becomes equal to s_{sit}, the actual sectoral share in county i and time t. Hence:

$$s_{sit} - \vartheta_{sit} = \\ \frac{\exp(\alpha_{si}+\beta_{1si}t+\beta_{2si}t^2+\beta_{3s})}{1+\exp(\alpha_{si}+\beta_{1si}t+\beta_{2si}t^2+\beta_{3s})} - \frac{\exp(\alpha_{si}+\beta_{1si}t+\beta_{2si}t^2+\beta_{3s}\delta_{it})}{1+\exp(\alpha_{si}+\beta_{1si}t+\beta_{2si}t^2+\beta_{3s}\delta_{it})} \tag{A5}$$

Combining equations (A3) and (A5) and solving for δ_{sit} yields:

$$\delta_{sit} = \delta_{it}\left[1 - \frac{\beta_{3s}}{(\alpha_{si}+\beta_{1si}t+\beta_{2si}t^2+\beta_{3s})}(1-\delta_{it})\right]. \tag{A6}$$

One difficulty with using regression equation (A4) in this setting is that the fitted values ought to sum up to one. While multivariate extensions of Papke and Wooldridge's (1996) generalised linear model for fractional response variables are available (e.g., Buis 2010), in practice estimation becomes challenging: with our specification it was not possible for the iterative procedure to converge towards the maximum likelihood estimator with Buis' (2010) method. A viable alternative is to constraint the marginal effects to (approximately) sum up to 0. The derivative with respect to each variable is equal to its coefficient times $exp(z)/[1 + exp(z)]^2$, where $exp(z)/[1 + exp(z)]$ is the fitted value. At the sample mean of the dependent variable, which by definition is equal to one-third, $exp(z)/[1 + exp(z)]^2 = 0.222$ for all sectors. Therefore at this value the condition that the marginal effects of time and share of deaths covered cancel themselves out across sectors simplify into:

$$\beta_{13i} = -\beta_{11i} - 2\beta_{21i}t - \beta_{12i} - 2\beta_{22i}t - 2\beta_{23i}t, \tag{A7}$$

and:

$$\beta_{33} = -\beta_{31} - \beta_{32} \tag{A8}$$

Substituting these conditions into (A4) for $s = 3$ and re-arranging yields:

$$\theta_{3it} = exp[\alpha_{3i} + \beta_{11i} (-t) + \beta_{12i} (-t) + \beta_{21i} (-2\, t^2) + \beta_{22i}(-2\, t^2) + \beta_{23i}(-t^2) + \beta_{31}(-\delta_{it}) + \beta_{32}(-\delta_{it})]/\{1 + exp[\alpha_{3i} + \beta_{11i} (-t) + \beta_{12i} (-t) + \beta_{21i} (-2\, t^2) + \beta_{22i}(-2\, t^2) + \beta_{23i}(-t^2) + \beta_{31}(-\delta_{it}) + \beta_{32}(-\delta_{it})]\} + u_{3it}. \quad \text{(A9)}$$

The first column of Table S2.2 shows the baseline specification; the second one allows the coefficient of the share of deaths covered (β_{3s}) to vary across decades; the third one only include data up to the 1750s. It is straightforward to compute the key coefficient for services with equation (A8) and for reasons of space its values are not presented here.

TABLE S2.2
SHARES OF DEATHS COVERED AND SECTORAL SHARES IN THE PROBATE RECORD: GENERALISED LINEAR REGRESSION ANALYSIS FOR FRACTIONAL RESPONSE VARIABLES

Sector	Period	(1)	(2)	(3)
Agriculture	1540–1809	0.586 (6.01)***		
	1540–1759			0.492 (5.96)***
	1540–1549		–0.430 (–1.38)	
	1550–1559		0.594 (7.16)***	
	1560–1569		0.513 (3.23)***	
	1570–1579		0.556 (9.30)***	
	1580–1589		0.531 (73.51)***	
	1590–1599		0.894 (7.26)***	
	1600–1609		0.584 (7.34)***	
	1610–1619		0.392 (4.43)***	
	1620–1629		0.250 (4.08)***	
	1630–1639		0.016 (0.81)	
	1640–1649		–0.777 (–6.14)***	
	1650–1659		–0.101 (–3.12)***	
	1660–1669		–0.886 (–11.14)***	
	1670–1679		–0.877 (–10.78)***	
	1680–1689		–1.123 (–10.68)***	
	1690–1699		–1.177 (–7.85)***	
	1700–1709		–1.258 (–8.24)***	
	1710–1719		–0.592 (–10.66)***	
	1720–1729		0.171 (2.47)**	
	1730–1739		–0.254 (–26.28)***	

Table S2.2–continued

	1740–1749		–0.802 (–20.32)***	
	1750–1759		0.179 (2.07)**	
	1760–1769		1.577 (6.00)***	
	1770–1779		3.213 (7.54)***	
	1780–1789		5.255 (7.69)***	
	1790–1799		6.788 (7.63)***	
	1800–1809		6.710 (7.74)***	
Industry	1550–1809	0.320 (2.53)**		
	1550–1759			0.254 (5.18)**
	1540–1549		–2.222 (–4.28)***	
	1550–1559		–1.324 (–8.69)***	
	1560–1569		–0.945 (–3.94)***	
	1570–1579		–0.507 (–5.52)***	
	1580–1589		–0.398 (–21.76)***	
	1590–1599		–0.252 (–1.91)*	
	1600–1609		–0.182 (–2.14)**	
	1610–1619		0.033 (0.34)	
	1620–1629		0.001 (0.01)	
	1630–1639		0.042 (1.71)*	
	1640–1649		0.087 (0.60)	
	1650–1659		–0.292 (–9.04)***	
	1660–1669		0.763 (8.87)***	
	1670–1679		0.735 (8.46)***	
	1680–1689		0.951 (8.62)***	
	1690–1699		0.579 (3.61)***	
	1700–1709		0.667 (4.22)***	
	1710–1719		0.638 (15.59)***	
	1720–1729		0.645 (6.16)***	
	1730–1739		0.860 (21.62)***	
	1740–1749		1.079 (77.08)***	

Table S2.2–continued

	1750–1759		1.091	
			(7.28)***	
	1760–1769		1.169	
			(3.24)***	
	1770–1779		0.192	
			(0.35)	
	1780–1789		–0.188	
			(–0.22)	
	1790–1799		–0.271	
			(–0.24)	
	1800–1809		–0.435	
			(–0.40)	
County/sector fixed effects		Yes	Yes	Yes
Quadratic county/sector trends		Yes	Yes	Yes
N		1809	1809	1452

* = Significant at the 1 percent level.
** = Significant at the 5 percent level.
*** = Significant at the 1 percent level.
Note: *N*=sample size. Clustered standard errors allow for arbitrary correlation within sectors; the *z-statistics* are in parentheses.
Sources: See Appendix 1 and 2 and the text.

REFERENCES

Brass, W. "On the Scale of Mortality." In *Biological Aspects of Demography*, edited by W. Brass, 69–110. London: Taylor and Francis, 1971.

Buis, L. M. 2010. "Relationship between the Likelihood Functions Used in fmlogit and mlogit." Available at http://www.maartenbuis.nl/software/likelihoodFmlogit.pdf.

Landers, J. *Death and the Metropolis. Studies in the Demographic History of London 1670–1830*. Cambridge: Cambridge University Press, 1993.

Papke, L. E., and J. M. Wooldridge. "Econometric Methods for Fractional Response Variables with an Application to 401(K) Plan Participation Rates." *Journal of Applied Econometrics* 11, no. 6 (1996): 619–32.

Woods, R. "Urban-rural mortality differentials: an unresolved debate". *Population and Development Review* 29, no. 1 (2003): 29-46.

Wrigley, E. A., and Schofield, R. S. *The Population History of England 1541–1871. A Reconstruction*. First paperback edition. Cambridge: Cambridge University Press, 1989.

Wrigley, E. A., Davies, R. S., Oeppen, J. E., et al. *English Population History from Family Reconstitution 1580–1837*. Cambridge: Cambridge University Press, 1997.