

Discussion Papers
Department of Economics
University of Copenhagen

No. 08-16

Malthus in Cointegration Space:
A new look at living standards and population
in pre-industrial England

Niels Framroze Møller
Paul Sharp

Stu­di­estræde 6, DK-1455 Copenhagen K., Denmark
Tel.: +45 35 32 30 82 – Fax: +45 35 32 30 00
<http://www.econ.ku.dk>

ISSN: 1601-2461 (online)

Malthus in Cointegration Space:

A new look at living standards and population in pre-industrial England*

Niels Framroze Møller and Paul Sharp

University of Copenhagen, Department of Economics

July 11, 2008

Abstract

We analyze Malthus' (1798) model when labor demand shifts persistently. The Malthusian ideas are formalized and derived in terms of stationarity and cointegration, and the implied restrictions are tested against English pre-industrial data 1560-1760. The evidence suggests a negligible marginal productivity effect of population on real income, implying that the Malthusian "check" relations should be analyzed as cointegrating relations. The data support highly significant preventive checks working via marriages, but weak (insignificant) positive checks. These results are remarkably clear-cut. We suggest a simple interpretation for the lack of response of real income to population, which is consistent with positive feed back effects from population on technology, à la Boserupian- and/or Smithian mechanisms. Recursive estimation confirms stable parameters and identify the end of our modified Malthusian regime.

JEL codes: C32, N3, O1.

Keywords: Cointegrated VAR, unit root econometrics, Malthus, Malthusian model, pre-industrial England.

*We thank Carl-Johan Dalgaard, Christian Groth, Hans Oluf Hansen, Ingrid Henriksen, Joannes Jacobsen, Søren Johansen, Katarina Juselius, Anders Bredahl Kock, Ronald Demos Lee, Diana Framroze Møller, Heino Bohn Nielsen, Karl Gunnar Persson, and Jacob Weisdorf. We also thank participants at the DGPE workshop 2007, participants at the 1st FRESH meeting at the Paris School of Economics, and Workshop participants from the Economics group at LUISS, Rome.

1 Introduction

Malthus' (1798) ideas of the interaction between population and the economy still occupy central stage in the economist's understanding of the pre-industrial economy. Malthusian assumptions are routinely invoked by growth theorists when modelling the pre-industrial world (see for example the survey by Galor 2005). Some have even gone so far as to claim that all of economic history before 1800 can be explained by the "Malthusian model" (Clark 2007).

The modern version of the Malthusian model explains the growth of real income and population. It involves *three core assumptions*: Higher labor income per capita increases the number of births (the *preventive check*), lowers the number of deaths (the *positive check*), and labor exhibits *diminishing marginal returns* due to the fixity of available land. According to the *classical* interpretation, higher income results from increasing labor demand which raises wage incomes (Lee 1997). This makes for population growth via the "checks", and eventually, implies a higher labor supply offsetting the initial wage increase cf. the diminishing returns. Hence, the core assumptions imply that labor demand at given real wages, is the driving force of population growth, and that individual living standards stay unaltered in the long run.

This paper shows how to analyze the Malthusian ideas when this driving force evolves *persistently*, which we shall model by a unit root process. Persistence seems a plausible assumption since the level of labor demand has usually been associated with the *accumulation* of technical knowledge (methods of cultivation and fertilization), and the level of capital per worker (see e.g. Lee 1997). By formalizing the persistence with a unit root I(1) process, the Malthusian ideas can be interpreted in terms of stationary and cointegrated Vector-Autoregressive (VAR) models. This enables a detailed understanding of the empirically observed persistence of income and vital rates based on the Malthusian framework, which has been absent in the literature.

Our approach also allows a systematic and explicit treatment of existing problems and issues in the literature, such as endogeneity of all variables (Lee 1973 and recently Nicolini 2007), dynamic persistence (Lee 1993a, for example), different orders of integration of interacting variables (Bailey and Chambers 1993)¹.

We confront our Malthusian model with the well-known English data from Wrigley and Schofield (1989) on births, marriages and deaths, and from Allen (2001), on real wages. For the period 1560-1760 the data suggest a negligible marginal productivity effect of population on the real wage rate. As a result, the real wage inherits the persistence of labor demand. This implies that the Malthusian check relations should be analyzed as cointegrating relations. The results support strongly significant preventive checks working via marriages, but weak (insignificant) positive checks. These conclusions are remarkably clear-cut, and recursive estimation confirms stable parameters.

We suggest a simple and tentative interpretation of the lacking response of real income to population, allowing for positive feedback effects from population on technology, i.e. supposedly Boserupian (and/or Smithian) mechanisms (See e.g. Simon 1977).

The next section presents a simple Malthusian model with persistent labor demand. Section

¹For a concise survey of such problems see Lee (1997) or Lee and Anderson (2002), p. 198.

2.2 elaborates the preventive check with the marriage rate. Section 2.3 shows that these models are equivalent to 1) Cointegrated VARs in population and labor demand (section 2.3.1), and 2) Stationary VARs in the observable real wage and vital rates (section 2.3.2). The necessary condition of stationarity is then tested against the English data, and rejected (section 3). Given the evidence of non-stationarity we analyze how the Malthusian model (of the real wage rate and vital rates) can generate such persistence (section 4.1). Two explanations exist: Either it is the sum of the check effects or the marginal productivity effect of population on real wages that is too weak. Both of these "persistence restrictions" are accepted by the data (section 4.2). We argue in favour of a negligible marginal productivity effect and our tentative interpretation of the evidence is suggested in section 5. The results are discussed in section 6, the robustness is analyzed in section 7, while section 8 concludes.

2 Formalization of the Malthusian ideas

Our first model is akin to the models in Lee (1993a), Lee (1997) and Lee and Anderson (2002), and corresponds to the traditional diagrams found in text books (e.g. Miller and Upton 1986, Clark 2007). The distinguishing feature is the persistence of labor demand shifts.

2.1 A simple Malthusian theory model

Our baseline Malthusian model is denoted by M_1 , with the equations,

$$w_t = c_0 - c_1 \ln N_t + \ln A_t, \quad (1)$$

$$b_t = a_0 + a_1 w_t + \varepsilon_{bt}, \quad (2)$$

$$d_t = a_2 - a_3 w_t + \varepsilon_{dt}, \quad (3)$$

$$\ln A_t = \ln A_{t-1} + \varepsilon_{At}, \quad (4)$$

$$\ln N_t = \ln N_{t-1} + b_{t-1} - d_{t-1}, \quad (5)$$

where w_t is the natural logarithm of the real wage, b_t , the crude birth rate, d_t , the crude death rate, and N_t is total population. In general, A_t comprises all determinants of labor demand at given real wages². All parameters are stated as positive. The shocks, ε_{bt} and ε_{dt} , represent unmodelled unsystematic influences on births and deaths respectively.

To give substance to M_1 , we can imagine a simple classical economy which is closed with respect to both trade and migration. It has a household sector and a production sector, interacting in markets for output and inputs. All prices are fully flexible and output is supply determined. As is typical when modeling pre-industrial economies it is assumed that the supply of available land is fixed (Lee 1973, p. 587, Clark 2007, p. 24, Galor 2005). Aggregate labor supply is assumed to be proportional to total population.

²We allow for a constant in (4), i.e. a deterministic linear trend in $\ln A$ in the empirical analysis (section 6).

Equation (1) describes the real wage at labor market equilibrium. The aggregate labor demand schedule is downward sloping due to diminishing marginal returns ($c_1 > 0$), and shifts when A_t changes. The evolution of A_t is modelled as the simple I(1) unit root process - the random walk in (4). This is the distinguishing feature of our exposition that captures the persistence in labor demand shifts.

Population, and hence, labor supply, evolve according to (5). Via $b_{t-1} - d_{t-1}$ it is determined by economic conditions, i.e. the preventive check, a_1 , in (2), and the positive check, $-a_3$ in (3). There are many explanations of a positive effect of real wages on births, ranging from biological to institutional circumstances (See e.g. Lee 1977, the introduction). We shall focus on the popular one based on marriages: At higher real wages more people marry, resulting in more births (section 2.2). The positive check is usually thought to work through nutrition, infant mortality, etc. (See e.g. Lee 1997, p. 1065, Schultz 1981).

If the production function is the constant-returns-to-scale Cobb-Douglas with capital, land and labor as inputs, c_1 in equation (1), will be the sum of the income shares to capital and land. In this case the real wage is proportional to real income per capita, and hence, both can be used as determinants of population growth, $b - d$.

This makes up the static equilibrium of M_1 . More importantly, the Malthusian economy also has a steady state, defined by zero population growth, towards which it converges in the absence of disturbances (ε shocks). With respect to the analysis of cointegration and stationarity below it is illuminating to consider this deterministic steady state.

When no shocks occur and A is fixed at \bar{A} , M_1 implies the steady state values;

$$w^* = \frac{a_2 - a_0}{a_1 + a_3}, \quad b^* = d^* = \frac{a_0 a_3 + a_1 a_2}{a_1 + a_3}, \quad \ln A^* = \ln \bar{A}, \quad \ln N^* = \frac{1}{c_1} (c_0 + \ln \bar{A} - \frac{a_2 - a_0}{a_1 + a_3}), \quad (6)$$

where, as is usual, w^* is referred to as the *subsistence* level³.

In M_1 , the three Malthusian core assumptions are that $c_1 > 0$ (diminishing marginal productivity), $a_1 > 0$ and $a_3 > 0$ (the check mechanisms). From (6) we see that these three assumptions ensure *existence* of steady state. This steady state is *stable* when $0 < c_1(a_1 + a_3) < 2$. Oscillations occur when $1 < c_1(a_1 + a_3) < 2$, which we define as "Malthusian Oscillations"⁴.

The model is illustrated in Figure 1. Graphically the shocks ε_{bt} and ε_{dt} correspond to stochastic intercepts of the two graphs in the left panel, and ε_{At} , to shifts in the downward sloping labor demand curve (right panel).

For later, we note that the steady state comparative static effect of a permanent unit rise in $\ln A^*$ on $\ln N^*$ is $\frac{1}{c_1}$, while this has no effect on w^* : A unit rise in $\ln A^*$ shifts the labor demand schedule to the right by $\frac{1}{c_1}$, and thus, in the long run the vertical labor supply schedule (population) must shift by this amount for the real wage to stay unaltered.

There are two ways to interpret the time horizon: The first, which we adopt in the theoretical

³This subsistence level may be above the minimal physiological requirements (Galor 2005, p. 179).

⁴Malthusian Oscillations have actually been claimed to be part of Malthus' core predictions (see e.g. Lee 1977, p. 347, Lee 1987, Lin Lee and Loschky 1987). We do not, however, find empirical support for their existence.

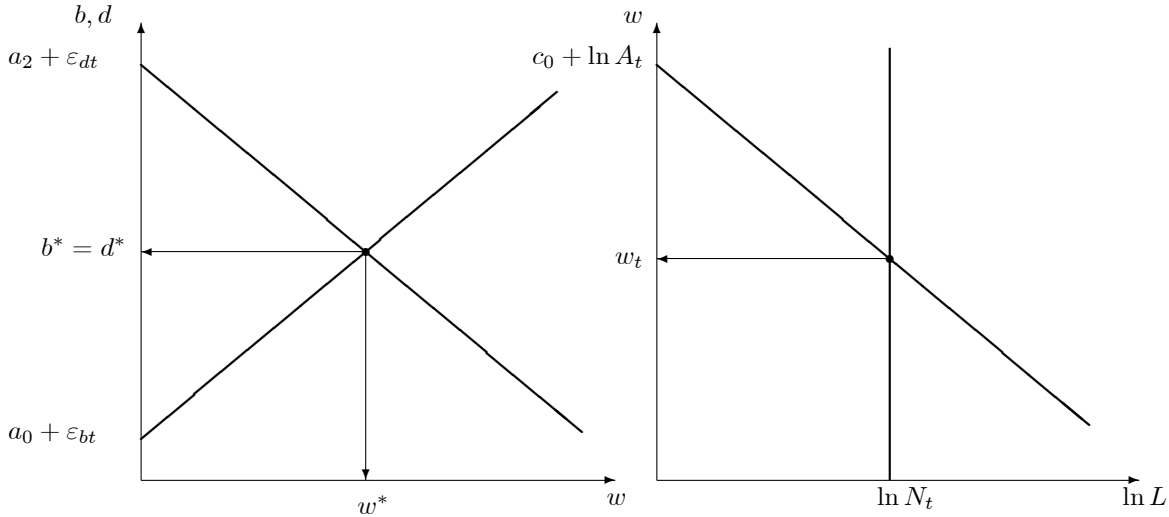


Figure 1: The simple Malthusian model, M_1 : The positive check relation (downward sloping) together with the preventive check relation (upward sloping) in the left panel. The downward sloping labour demand and vertical labour supply in the right panel.

analysis, is to choose the length of period to make the theoretical assumptions defensible. For example, births in period t should only matter for the real wage in period $t + 1$ if the periods are 10-20 years long, i.e. the time it takes to enter the work force. However, given yearly data (section 3.2), it seems a tremendous loss of information to aggregate into periods of 10-20 years. Hence, the second possibility is to take the length of the period as given and instead allow for gradual dynamic adjustment by adding lags. This is the approach in the empirical analysis.

2.2 Adding the marriage rate

It has been widely believed that a considerable part of the preventive check worked through marriages, also by Malthus himself (Malthus (1798), Lee 1997 p. 1065). We therefore augment M_1 with the crude marriage rate, m_t . We denote the resulting model M_2 , which is simply M_1 with (2) replaced by (9) and (14) below.

From the breadwinner's point of view, marriage implied provision of spouse and children for several years, and hence, expected future income was a crucial determinant in the marriage decision. Here, we shall assume that income expectations are adaptive, and of the form,

$$w_{t+i}^e = w_t, \quad (7)$$

where superscript e denotes the (subjective) expectation. The assumption in (7) seems reasonable in light of the observed persistence of the real wage (see section 3.3), and provides a simple and manageable point of departure. An alternative could be Rational Expectations, or Model Consistent Expectations in the sense of Muth (1961). As this implies that agents form expectations about the interaction between income and vital rates, we believe that this is not a natural point of departure. However, when the real wage rate is a random walk, which is accepted empirically (see section 3.3), the two types of expectations formation coincide.

We assume that the share of marriages in the population, m_t , depends positively on the expected future income profile, i.e. that,

$$m_t = \beta_{01} + \beta_1^1 w_t + \beta_2^1 w_{t+1}^e \dots + \beta_j^1 w_{t+j-1}^e + \varepsilon_{mt}. \quad (8)$$

The coefficients, $\beta_1^1, \dots, \beta_j^1$ are positive and capture some kind of discounting. The scalar, j , can be interpreted as life expectancy at the age of marriage, or perhaps more likely, as a shorter planning horizon. The error term, ε_{mt} , captures unmodelled unsystematic influences.

From (7) and (8) it follows that,

$$m_t = a_4 + a_5 w_t + \varepsilon_{mt}, \quad (9)$$

where a_5 thus depends on j and β_i^1 ($a_5 > 0$). Note that, if the real wage rate has a linear deterministic trend this will be captured by a_4 .

The coefficient, a_5 , in (9) thus describes part of the preventive check effect, i.e. from income to marriages. The rest depends on how marriages affect births.

Bailey and Chambers argue that, though tempting, one should not include the marriage rate in a birth rate relation, as the relevant variable is the *stock* of *fertile* marriages, and not the *flow* measure, m_t (Bailey and Chambers 1993, p. 346). Since data of m_t are available only, they choose not to model the effect from marriages on births. We agree that the relevant measure is the stock of fertile marriages. However, it *is* possible to relate the birth rate, b , to the marriage rate, m , as the stock of fertile marriages depends on the sum of marriages over say, the last s years, and hence, on the marriage rates, m_{t-s}, \dots, m_{t-1} :

We define a *fertile marriage* as a marriage in which the woman is in her fertile age (the age between menarche and menopause). The stock of fertile marriages at the beginning of period t is denoted by M_t^f . The total number of births (apart from an error term) in period t , B_t , is assumed to be proportional to this, i.e.,

$$B_t = \zeta M_t^f, \quad \zeta > 0, \quad (10)$$

abstracting from illegitimacy, which was probably negligible (3-4%, according to Clark 2007).

Assume now that the age distribution in period t , of women married in period t , is approximately independent of t . If the number of new marriages during period $t - i$ is M_{t-i} (a flow measure), and the minimum age of the marrying women is x , we can write,

$$M_t^f = \delta_1 M_{t-1} + \delta_2 M_{t-2} + \dots + \delta_s M_{t-s}, \quad (11)$$

where $s = z - x$, z being the (average) age at menopause. The weights, $\delta_i > 0$, decline for two reasons: First, the effect from M_{t-i} on M_t^f will be smaller the greater i is, as fewer of the women married in period $t - i$ will be in their fertile age in period t . Second, some "discounting" due to deaths, divorce, declining fecundity etc., also seems natural. The decline in weights implies that empirically we can probably approximate the dynamics with less than s lags.

Divide (11) by N_t , and use that $\frac{M_{t-i}}{N_t} = \frac{N_{t-i}}{N_t} \frac{M_{t-i}}{N_{t-i}} = \frac{1}{(1+n_{t-1}) \dots (1+n_{t-i})} m_{t-i}$, to obtain,

$$\frac{M_t^f}{N_t} = \frac{\delta_1}{1+n_{t-1}} m_{t-1} + \frac{\delta_2}{(1+n_{t-1})(1+n_{t-2})} m_{t-2} + \dots + \frac{\delta_s}{(1+n_{t-1}) \dots (1+n_{t-s})} m_{t-s}, \quad (12)$$

where n_{t-i} is the population growth rate, equal to $b_{t-i} - d_{t-i}$. It now follows directly from (10) and (12) that our birth rate relation (plus a disturbance) becomes,

$$b_t = \frac{\zeta \delta_1}{1+b_{t-1}-d_{t-1}} m_{t-1} + \dots + \frac{\zeta \delta_s}{(1+b_{t-1}-d_{t-1}) \dots (1+b_{t-s}-d_{t-s})} m_{t-s} + \varepsilon_{bt}. \quad (13)$$

This is a non-linear difference equation of order s . For the empirical analysis below, we thus have two possibilities: Either we can use (13) directly in the econometric model, or we can make a linear approximation of it, which fits into the usual framework of the VAR.

A linear approximation requires some point of (b, m, d) to approximate around. Under stability of the deterministic steady state, w_t, b_t, d_t and m_t are stationary, and their steady state values (means) can be used to approximate around. Using that $b^* = d^*$ or $n^* = 0$, we find, from a first order Taylor approximation of (13) around steady state (b^*, d^*, m^*) , or equivalently, around (n^*, m^*) , that,

$$b_t \simeq e_0 + e_1 b_{t-1} - e_1 d_{t-1} + f_1 m_{t-1} + e_2 b_{t-2} - e_2 d_{t-2} + f_2 m_{t-2} + \dots + e_s b_{t-s} - e_s d_{t-s} + f_s m_{t-s} + \varepsilon_{bt}, \quad (14)$$

where $e_0 \equiv b^* + e_1$, $e_l \equiv -m^* \zeta \sum_{i=l}^s \delta_i$ and $f_l = \zeta \delta_l$ for $l = 1, 2, \dots, s$.

Given existence of the stable steady state, the empirical adequacy of (14) as an approximation of (13), will depend on the degree of non-linearity of (13) within the realistic range of data variation. If this is not too pronounced the approximation is useful. This is likely to be the case here: To see this, assume that $s = 1$ in (13). The expression to be approximated is then,

$$F(n_{t-1}, m_{t-1}) \equiv \frac{\zeta \delta_1}{1+n_{t-1}} m_{t-1}, \quad (15)$$

and the approximation, defined as $F^a(n_{t-1}, m_{t-1})$, becomes,

$$F^a(n_{t-1}, m_{t-1}) = -\frac{m^*}{(1+n^*)^2} \zeta \delta_1 (n_{t-1} - n^*) + \frac{\zeta \delta_1}{1+n^*} m_{t-1}. \quad (16)$$

First note from (15), that F is linear in m_{t-1} at a given n_{t-1} . Hence, the approximation is independent of the variation in m_{t-1} . In contrast, F is not linear in n_{t-1} given m_{t-1} , and thus, for deviations in the n_{t-1} -direction the adequacy of the approximation will depend on both the degree of non-linearity at (n^*, m^*) and the variation range around this. Since population is non-negative, $n > -1$ always, but for n close to -1 the slope of F changes very fast, and hence, the constant slope implied by the linear approximation is inappropriate. Fortunately, in the Malthusian equilibrium, $n^* = 0$, and the empirical variation of n_t is located within ± 0.02 . The

curvature in the n_{t-1} -direction also depends on m^* , but if the model is to have any empirical relevance this will probably lie within a limited range, 0.005-0.02 say, (the observed range of m_t is 0.005-0.015). When $s > 1$ the same arguments apply. Altogether, for the empirically relevant region of data variation the linear approximation seems adequate.

The approximation will still work if b_t and d_t are I(1), provided that $n_t = b_t - d_t$ is I(0), i.e. that b_t and d_t cointegrate. Also, if some (combination of) variables are "stationary" when corrected for level shifts, as is typical in applications, one can think of the linear approximation as being around the corrected mean. This will imply non-linear restrictions on the coefficients of the dummies in the estimation equation which we shall ignore.

Analogously to M_1 , the steady state of M_2 for $s \geq 1$ is given by (6) with a_0 and a_1 replaced by $a_0^* \equiv e_0 + fa_4$ and $a_1^* \equiv fa_5$ respectively, where $f \equiv f_1 + \dots f_s$. The steady state marriage rate is $m^* = a_4 + a_5 w^*$. As in M_1 existence is thus guaranteed by the core assumptions, $c_1 > 0$, $a_3 > 0$, and, in this case, $a_1^* = fa_5 > 0$.

2.3 Implications in terms of stationarity and cointegration

When labor demand is driven by persistent forces modeled by the random walk in (4), the Malthusian ideas can be formulated as M_1 , or M_2 when the marriage rate is included. We shall now show that each of these models are equivalent to, 1) a cointegrated VAR model in the stocks, $(\ln N_t, \ln A_t)'$, with one cointegrating relation, and $\ln A_t$ as the common stochastic trend, and 2) a stationary VAR model in the rates, (w_t, b_t, d_t) , or (w_t, b_t, d_t, m_t) ⁵. Corresponding to M_i , $i = 1, 2$, we denote the cointegrated VARs by M_i^S , and the stationary VARs by M_i^R . As A_t is unobserved only models in the rates are taken to the data (section 3.3).

The VAR(k) with a constant in Error-Correction-Mechanism (ECM) form is given by,

$$\Delta x_t = \Pi x_{t-1} + \Gamma_1 \Delta x_{t-1} + \dots + \Gamma_{k-1} \Delta x_{t-(k-1)} + \mu + \varepsilon_t, \quad (17)$$

where x_t is p -dimensional⁶. If the characteristic roots, z , are always either at 1 or outside the unit disc, then x_t is non-stationary when at least one root is at 1, or equivalently, when $\det(\Pi) = 0$. This implies reduced rank, $r < p$, of Π , which is parameterized as the non-linear restriction $\Pi = \alpha\beta'$. The matrices α and β are $p \times r$ of rank r , and for $0 < r < p$, r cointegration relations exist. Stationarity of x_t is thus equivalent to $\det(\Pi) \neq 0$ (Π has full rank).

The essential mechanisms can be studied based on M_1 alone. As M_2 does not change the long-run dynamics, this is treated briefly in continuation of M_1 .

⁵By stationary we generally mean "asymptotically stationary" (See Johansen 1996, p 15 for example).

⁶For technical details we generally refer to Johansen (1996).

2.3.1 The cointegrated VARs of population and labor demand, M_1^S and M_2^S

Solving M_1 with respect to $(\Delta \ln N_t, \Delta \ln A_t)'$ we immediately obtain M_1^S :

$$\begin{aligned}\Delta \ln N_t &= \mu_0 + \alpha_1(\ln N_{t-1} - \beta_1 \ln A_{t-1}) + \varepsilon_{Nt}, \\ \Delta \ln A_t &= \varepsilon_{At},\end{aligned}\tag{18}$$

where $\alpha_1 \equiv -c_1(a_1 + a_3)$, $\beta_1 \equiv \frac{1}{c_1}$, $\mu_0 \equiv (a_0 - a_2) + c_0(a_1 + a_3)$ and $\varepsilon_{Nt} \equiv \varepsilon_{bt-1} - \varepsilon_{dt-1}$. With $x_t = (\ln N_t, \ln A_t)'$ this involves the parameters,

$$M_1^S : \Pi = \begin{pmatrix} \alpha_1 & -\alpha_1\beta_1 \\ 0 & 0 \end{pmatrix}, \alpha = \begin{pmatrix} \alpha_1 \\ 0 \end{pmatrix}, \beta = \begin{pmatrix} 1 \\ -\beta_1 \end{pmatrix}.\tag{19}$$

The roots are $z_1 = 1$, and $z_2 = \frac{1}{1+\alpha_1}$ for $\alpha_1 \neq -1$, and $z = 1$ for $\alpha_1 = -1$. If the steady state in (6) is stable, it follows that $|z_2| > 1$, where $|\cdot|$ denotes the modulus. Given this, and since $\det(\alpha'_\perp \beta_\perp) \neq 0$, where α_\perp and β_\perp are the orthogonal complements, $(\ln N_t, \ln A_t)'$ will be I(1) with cointegrating vector $(1, -\beta_1)'$. See theorem 4.2 in Johansen (1996).

The adjustment vector in M_1^S is $\alpha' = (\alpha_1, 0)$, which implies that $\ln A_t$ is strongly exogenous (see Johansen 1992), and that $\sum_{i=1}^t \varepsilon_{Ai}$ is the common stochastic trend. This captures that labor demand drives population. Since μ is proportional to α , x_t contains no linear deterministic trend. Equivalently, adding an intercept in the equation for $\ln A_t$ introduces such a trend in both variables.

The long-run impact matrix, $C = \beta_\perp(\alpha'_\perp \beta_\perp)^{-1} \alpha'_\perp$, is,

$$C = \begin{pmatrix} 0 & \beta_1 \\ 0 & 1 \end{pmatrix},\tag{20}$$

which shows that the long-run impact of a unit shock to labor demand, i.e. $\varepsilon_{At} = 1$, on $\ln N$ is $\beta_1 = \frac{1}{c_1}$. This is the steady state comparative static effect computed from (6). The transition period in the wake of such a shock, i.e. the sequences of "short run" equilibria, may be brief or prolonged, depending on whether α_1 is "large" or "small" respectively.

The C matrix also shows that there is no long-run impact on population from ε_{bt} and ε_{dt} : For example, a positive population - or labor supply shock lowers real wages which reduces population as long as $w < w^*$.

It is important to note that the unit root, i.e. $z_1 = 1$, results solely because $\ln A$ in equation (4) is a unit root process. If (4) were to be replaced by, say the stationary AR(1) process, $\ln A_t = \rho \ln A_{t-1} + \varepsilon_{At}$, z_1 would equal $\frac{1}{\rho}$ (z_2 unchanged). If A is interpreted as the level of technical knowledge, $\rho = 1$ would be a natural assumption. As a result, $z_1 = 1$, which would thus be an example of a *structural unit root*. On the other hand, A is more likely to comprise many other determinants, for which ρ is probably below 1, but supposedly close to 1, i.e. $\ln A$ is persistent. In that case, $z_1 = 1$ (or $\rho = 1$) is viewed as a useful *statistical approximation*, in the sense of delivering more reliable inference given the data at hand, than that based on

stationarity (see e.g. Møller 2008).

To derive M_2^S we assume that $s = 1$. In terms of long-run dynamics there is no substantial difference when $s > 1$ (Appendix A.1). In light of the analysis of M_1 we proceed in two steps:

First, we insert (9) into (14) to obtain a modified birth relation,

$$b_t = \check{a}_0 + e_1(b_{t-1} - d_{t-1}) + \check{a}_1 w_{t-1} + \check{\varepsilon}_{bt}, \quad (21)$$

where $\check{a}_0 \equiv e_0 + f_1 a_4$, $\check{a}_1 \equiv f_1 a_5 > 0$ and $\check{\varepsilon}_{bt} \equiv \varepsilon_{bt} + f_1 \varepsilon_{m_{t-1}}$. Equation (21) can be compared to (2). First, note that it is w_{t-1} that enters (21), and not w_t as in (2). This difference is however not substantial, and results since b_t depends on the lagged marriage rate, m_{t-1} . In fact, with "long periods" ($s = 1$) we should really include m_t in (14), but we ignore this here. The appearance of d_{t-1} in the birth relation should not be confused with a direct structural effect from deaths to births, such as that modelled elsewhere in the literature (e.g. Lee 1985b or Bailey and Chambers 1993). It follows purely from our derivations above. The coefficient, \check{a}_1 , clearly illustrates that the preventive check is the product of the two effects: higher income means more marriages, $a_5 > 0$, and more marriages mean more births, $f_1 > 0$. For $s > 1$ \check{a}_0 and \check{a}_1 are denoted a_0^* and a_1^* , respectively (section 2.2 and Appendix A.1).

As (21) replaces (2) in M_1 , the next step is to solve the resulting model for $(\ln N_t, \ln A_t)'$,

$$\begin{aligned} \Delta \ln N_t &= \check{\mu}_0 + \check{\alpha}_1 (\ln N_{t-1} - \beta_1 \ln A_{t-1}) + \theta_1 \Delta \ln N_{t-1} + \theta_2 \Delta \ln A_{t-1} + \check{\varepsilon}_{Nt}, \\ \Delta \ln A_t &= \varepsilon_{At}, \end{aligned} \quad (22)$$

where $\check{\mu}_0 \equiv (\check{a}_0 - a_2) + c_0(\check{a}_1 + a_3)$, $\check{\alpha}_1 \equiv -c_1(\check{a}_1 + a_3)$, $\theta_1 \equiv (e_1 + \check{a}_1 c_1)$, $\theta_2 \equiv -\check{a}_1$ and $\check{\varepsilon}_{Nt} \equiv \check{\varepsilon}_{bt-1} - \varepsilon_{dt-1}$. Compared to (18), the only difference in the model form is the additional lagged differences in the population equation. The long-run dynamics are unaltered, and labor demand still drives population. The adjustment of population has just become more complicated. As in M_1 there is a root at 1, and if the corresponding deterministic steady state is stable, the rest of the roots will have moduli greater than 1. Since $\det(\alpha'_\perp \Gamma \beta_\perp) \neq 0$, this implies that $(\ln N_t, \ln A_t)'$ is I(1).

This section should thus be clarifying relative to previous (verbal) discussions of cointegration between population and labor demand (Lee and Anderson (2002), p. 201).

Since $\ln A_t$ is unobserved, M_1^S and M_2^S only serve analytical purposes. For empirical purposes we need models for the observables.

2.3.2 The stationary VARs of the real wage and vital rates, M_1^R and M_2^R

Consider first M_1^R , which is formulated in (w_t, b_t, d_t) . By inserting (5) and (4) in (1), and using that $w_{t-1} = c_0 - c_1 \ln N_{t-1} + \ln A_{t-1}$, we eliminate the state variables, $\ln N_t$ and $\ln A_t$. Rewriting

(2) and (3), and solving M_1 with respect to $(\Delta w_t, \Delta b_t, \Delta d_t)'$ we obtain,

$$M_1^R : \Pi = \begin{pmatrix} 0 & -c_1 & c_1 \\ a_1 & -(1 + a_1 c_1) & a_1 c_1 \\ -a_3 & a_3 c_1 & -(1 + a_3 c_1) \end{pmatrix}, \quad \det(\Pi) = -c_1 (a_1 + a_3) \equiv \alpha_1 < 0. \quad (23)$$

As $z = \frac{1}{1+\alpha_1}$, the (w_t, b_t, d_t) -process is thus stationary given stability of the steady state.

Turn to M_2^R . With $s = 1$ we obtain,

$$M_2^R : \Pi = \begin{pmatrix} 0 & -c_1 & c_1 & 0 \\ 0 & -(1 - e_1) & -e_1 & f_1 \\ -a_3 & a_3 c_1 & -(1 + a_3 c_1) & 0 \\ a_5 & -a_5 c_1 & a_5 c_1 & -1 \end{pmatrix}, \quad \det(\Pi) = c_1(\check{a}_1 + a_3) \equiv -\check{\alpha}_1 > 0. \quad (24)$$

Again, this implies that the (w_t, b_t, d_t, m_t) -process is stationary.

Completely analogously, when $s > 1$ we get a Π matrix as in (24) where e_1 is replaced by $\bar{e}_1 \equiv \sum_{i=1}^s e_i$, and f_1 replaced by $f = \sum_{j=1}^s f_j$. The determinant is then $-\alpha_1^* > 0$, with $\alpha_1^* \equiv -c_1(a_1^* + a_3)$ and $a_1^* = a_5 f$ (Appendix A.1).

To sum up, the Malthusian model with a stable deterministic steady state and labor demand evolving persistently, implies 1) that labor demand cointegrates with population - the former driving the latter, and 2) that real wages and vital rates interact in a stationary VAR. It is the three Malthusian core assumptions alone that imply this.

Stationarity of the VAR in the rates is thus a necessary and testable condition of the Malthusian model, and thus, it seems sensible to test this before elaborating the analysis.

3 Testing the Malthusian theory

3.1 Empirical implementation

In order to illustrate the central points as clearly as possible, the Malthusian models so far have been highly simplified. As a result they are too abstract to take to the data directly. A more general Malthusian model which allows for more flexible dynamics of adjustment but has the same long-run properties is presented in Appendix A.2. This model is a generalization of M_2^R , and is denoted \tilde{M}_2^R . Compared to M_2^R it allows the current marriage rate (in addition to the lagged rates) to enter the birth rate equation, as well as other effects from the lagged b and d (in addition to the e_i coefficients in (14)). The death - and marriage rate equations now allow for more gradual adjustment in response to changes in w . The real wage equation is not generalized, but rather the empirical consequences of such generalizations are discussed in light of the evidence (section 6).

The first point to note here is that the previous results generalize: In particular, stationarity is still the result of the core assumptions. To see this we simply compute the determinant of

the respective Π matrix. This is done in (52) in Appendix A.2, restated here:

$$\det(\Pi) = c_1 (\tilde{a}_1 + \tilde{a}_3) (1 - \lambda)(1 - \delta)(1 - \gamma - \phi), \quad (25)$$

provided that $(1 - \lambda) \neq 0$, $(1 - \delta) \neq 0$, $(1 - \gamma - \phi) \neq 0$, and where,

$$\tilde{a}_1 \equiv \tilde{a}_5 \tilde{f}, \tilde{f} \equiv \frac{\sum_{i=0}^k f_i}{(1 - \gamma - \phi)}, \tilde{a}_3 \equiv \frac{-\psi}{1 - \lambda}, \tilde{a}_5 \equiv \frac{\eta}{(1 - \delta)}. \quad (26)$$

The definitions of the remaining parameters are found in Appendix A.2. The special case M_2^R is obtained by setting: $\gamma_i = e_i$, $\phi_i = -\gamma_i$, $f_0 = 0$ and $k = s$ in (57), $\psi_0 = -a_3$, $\psi_i = 0$ and $\lambda_i = 0$, for $i = 1, 2..k$, in (58), and $\eta_0 = a_5$, $\eta_i = 0$ and $\delta_i = 0$ for $i = 1, 2..k$, in (59).

The main point to note is that the " $\tilde{\cdot}$ " - parameters, i.e. the long-run parameters, are the general Malthusian parameters. Thus, in this more general model, stationarity, i.e. $\det(\Pi) \neq 0$, is still the result of the core assumptions implying that $c_1(\tilde{a}_1 + \tilde{a}_3) > 0$.

Given this more general model, \tilde{M}_2^R , and that potentially, many other short-run interactions may take place without affecting the long-run structure (see e.g. Lee 1977), the initial statistical model should allow for an unrestricted lag length - and structure. Our statistical model is thus the unrestricted VAR,

$$\Delta x_t = \Pi x_{t-1} + \Gamma_1 \Delta x_{t-1} + \dots + \Gamma_{k-1} \Delta x_{t-(k-1)} + \Phi \mathcal{D}_t + e_t, \quad (27)$$

with $e_t \sim Niid(0, \Omega)$, $\Phi \mathcal{D}_t$ being drift terms included to allow for deterministic growth in capital and technology, and/or dummy terms to condition on warfare and plague etc.

Provided that (27) is *statistically adequate* we can infer on \tilde{M}_2^R as this is nested. More importantly, in case the theory model rejects, the estimate of (27) offers clues as to how to modify and improve the theory.

3.2 On the data

To test the Malthusian model we use the well-known yearly data on the birth rate, the marriage rate and the death rate from Wrigley and Schofield (1989). These are all crude rates and defined as the number of births, marriages and deaths per one thousand head of population. These series are constructed using the so-called back projection method. Although this method is subject to certain criticisms (Lee 1985a, Lee 1993b), these are the only data available.

As the real income series, we use the real wage series for building laborers in London from Allen (2001). Although Clark (2005) presents a new national series, we use Allen's series to facilitate comparison to the results in Nicolini (2007). However, as long as the long-run fluctuations are roughly the same the choice of real wage series is of little significance. The data series are plotted in Figure 2. The graphs seem to suggest that the levels can be described with I(1) *stochastic* trends, since their first differences (growth rates) seem stationary, while linear *deterministic* trends seem absent. Both suggestions are supported by the tests below.

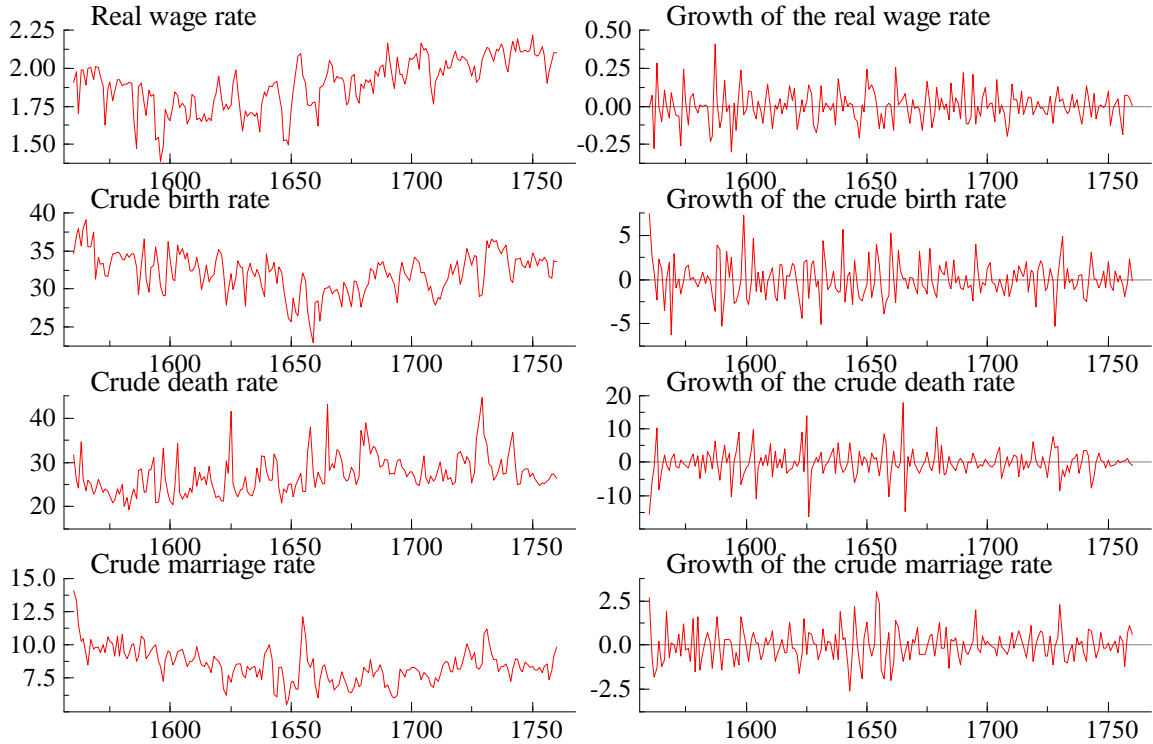


Figure 2: The graphs of the time series: Levels in the left panel, growth rates in the right panel.

We restrict our sample to the pre-industrial period, 1560-1760, which is usually regarded as safely within the period associated with the Malthusian regime. In section 7 we extend the sample to 1850 to get some idea of the impact of the Industrial Revolution on our estimates. Although data are available from 1541, we start in 1560, as the period 1541-1559 was too unstable to model satisfactory, probably due to warfare and religious conflict.

3.3 Testing stationarity

All estimation is calculated with CATS in RATS (Dennis, Hansen, and Juselius 2006). The initial statistical models have the form as in (27), with ΦD_t including a constant and a trend (CIDRIFT in CATS). To take account of extraordinary exogenous events (warfare and plague), we start with a fairly general formulation. In particular, CATS estimates the model,

$$\Delta x_t = \Pi x_{t-1} + \lambda D_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta x_{t-i} + \phi \Delta D_t + \phi_1 \Delta D_{t-1} + \dots + \phi_{k-1} \Delta D_{t-(k-1)} + e_t, \quad (28)$$

(ignoring the constant and trend) where $x'_t = (w_t, b_t, d_t, m_t)$ and D_t is a vector of shift-dummies. This model allows for both transitory and permanent effects from extraordinary events.

Under the non-linear restrictions $\lambda = -\Pi\phi$ and $\phi_i = -\Gamma_i\phi$, $i = 1, \dots, k$, this is the so-called additive -, or "level shift" model. In such a case the observed series, x_t , results from superimposing (or adding) ϕD_t on to an underlying VAR. Essentially, the latter is simply shifted by ϕ from some date and onwards. These non-linear restrictions are ignored in CATS.

In the unit root case when $\Pi = \alpha\beta'$, the non-linear restriction which is ignored is now $\beta'\phi$ in $\lambda = -\alpha\beta'\phi$, and hence the parameterization $\delta = -\beta'\phi$ is used, so that $\lambda = \alpha\delta$, which fits directly into the usual reduced rank algorithm (Johansen 1996, Chapter 6): That is, $\Pi = \alpha\beta'$, and so $\Pi x_{t-1} + \lambda D_{t-1} = \alpha\beta'x_{t-1} + \alpha\delta D_{t-1} = \alpha(\beta', \delta)(x_{t-1}, D_{t-1})' = \alpha\beta^*x_{t-1}^*$. By treating the level of D_t like this, *similarity* in the unit root - or rank test is obtained, as the effect on x_t from the deterministic term is the same under the null and the alternative. This results from the "additive type" of nature. Likewise the "CIDRIFT" specification of the trend yields similarity.

To begin with, we estimated a model with no dummies, and $k = 2^7$. As expected, both from analyzing demographic data in general, and the discussion in section 2.2, lag tests suggested more lags. So henceforth, all specifications have $k = 3$ or 4.

A model with no dummies and 4 lags seemed statistically adequate. We removed Δw_{t-3} and Δm_{t-3} as they were insignificant. This specification is denoted S_1 . The misspecification tests are given in Appendix B.1: The assumption of no autocorrelation is clearly the most important one, since inconsistency of the estimators result otherwise. Fortunately, in all specifications autocorrelated errors were convincingly rejected. Joint normality was however rejected, primarily due to the death rate residuals, particularly in the years 1625 and 1665 (respectively 4.5 and 5.39). These were obviously due to outbreaks of the plague. For the same reasons there was a moderate residual (2.9) in 1603. LM tests for multivariate ARCH were not accepted - a conclusion for all specifications. However, cointegrating inference is relatively robust towards ARCH, in contrast to autocorrelation (Rahbek, Hansen, and Dennis 2002).

A specification with one "impulse type" dummy denoted D_t^P , which is 1 in the three years 1603, 1625 and 1665 and zero otherwise, and enters as in (28), seemed to capture the impact of the plague. We call this S_2 . The estimates of ϕ_1 , ϕ_2 and ϕ_3 were insignificant and thus restricted to zero. From Appendix B.2 we see that, compared to S_1 , joint normality, though still rejected, now improved considerably (the test statistic dropped from 81.91 to 36.22). Constancy of parameters is analyzed in Section 7, when the rank has been determined.

S_2 thus seemed reasonable to continue with, and there were no notable outliers left. However, as is well known residuals from OLS (by which the unrestricted VAR is estimated) may hide large underlying errors due to the minimization of *squared* residuals. Thus, there may be exogenous events with enormous influence on the estimation, which are difficult to identify from the residuals. Therefore, with permission from the author, we used the influence diagnostics based on the *displacements* of the likelihood and the eigenvalues (Nielsen 2008). Though tentative, some results were quite clear: First of all, the three years of plague, 1603, 1625 and 1665 stuck out, as single influential observations. The year 1659 seemed to influence the estimated eigenvalues and hence, potentially unit root inference. The same was true for the period 1643-44, which is likely to be the result of the First English Civil War (1642-46). In general, the mid seventeenth century was a turbulent period, with three civil wars during 1642-1651, succeeded by The First Anglo-Dutch war (1652-54) and the Anglo-Spanish War (1654-60). It is notable that the residuals for the years 1603, 1643 and 1659 were all only moderately large

⁷The computer - output and code from CATS in addition to that given here can be obtained from the authors.

(2.9 for 1603 and below 3.2 for 1643 and 1659), and hence, would probably not have been detected from the residuals.

This led to our preferred specification S_3 , which is S_2 augmented with the shift-dummies, $D_{S_t}^{1643}$ and $D_{S_t}^{1659}$ entering as in (28).

In S_3 we tested stationarity, i.e. the rank of Π , as there were no indications of roots inside the unit disc - a result in all specifications. As the results from the unit root test (trace test) were not sufficiently clear-cut, we considered the estimated characteristic roots, the graphs of the recursively calculated trace test, the significance of adjustment coefficients α_{ij} , and the graphs of the cointegrating relations, $\beta'x_t$ (see Juselius 2006, Chapter 8). To assess robustness of the choice of rank r , we considered all these pieces of information, for S_1 and S_2 with four lags, and S_3 with three and four lags. In S_1 and S_2 , $k = 3$ was rejected. The results are summarized in Table 1. The graphs of the recursive trace test, and the characteristic roots suggested the same for all specification, $r = 3$ or 4, and are therefore not included in the table.

Table 1: Inference on the Cointegration Rank

	Model specification			
	S_1	S_2	S_3	
	$k = 4$	$k = 4$	$k = 3$	$k = 4$
Trace test	$r = 3$	$r = 4$ (3*)	$r = 4$ (3**)	$r = 4$ (3*)
Graph of $\beta'x_t$	$r = 2$ or 3	$r = 3$	$r = 3$	$r = 2$ or 3
Significance of α	$r = 3$	$r = 3$ or 4	$r = 4$	$r = 4$

Notes: CIDRIFT was used. S_1 and S_2 exclude Δw_{t-3} and Δm_{t-3} . S_3 excludes Δw_{t-3} .

3* (3**) means that $r = 3$ is accepted at 97.5% (99%) level of significance.

In case of dummy variables, critical values for the trace test are simulated in CATS

When valid, the Bartlett corrected trace test is used instead.

The results are roughly invariant with respect to the specification. At first sight, it seems that doubt is centered on whether $r = 3$ or $r = 4$. However, given this inconclusive evidence, there are several reasons why we choose $r = 3$: First of all, the borderline case for $r = 4$ of the trace test (basically all specifications would have picked $r = 3$ on the 97.5% level), in conjunction with the graphs of $\beta'x_t$ suggest $r = 3$. Secondly, the data are inevitably subject to measurement errors which via their "additive outlier" nature may bias the trace test in favour of $r = 4$, i.e. stationarity (Franses and Haldrup 1994). Thirdly, at the stationary/non-stationary border it may be more appropriate to impose exact unit roots and use the asymptotics for unit root processes rather than for stationary processes (see e.g. Johansen 2006).

Altogether, it seems reasonable to conclude that Π has reduced rank, $r = 3$, or at least that imposing $r = 3$ is a good statistical approximation. Hence, in contrast to Nicolini (2007) we reject the stationarity ($r = 4$) implied by the Malthusian model. Note that we test stationarity as a system property, in contrast to Nicolini (2007) and more recently Crafts and Mills (2007) who adopt a univariate ADF-approach.

So, what explains this outcome? There are two possibilities; The Malthusian interactions are too sluggish for stationarity to be accepted. Alternatively, there is some other reason for persistence. As there may be many potential alternatives, it seems sensible first to investigate how the Malthusian model in the observable $x'_t = (w_t, b_t, d_t, m_t)$ can generate persistence.

4 Persistence in the Malthusian model

4.1 Understanding persistence in a Malthusian framework

As we have seen stationarity of $x'_t = (w_t, b_t, d_t, m_t)$ is equivalent to $\det(\Pi) \neq 0$. When $\det(\Pi) \rightarrow 0$ persistence emerges consistent with the evidence in section 3.3. We analyze this limit case by imposing $\det(\Pi) = 0$, equivalently the unit root restriction. This gives a statistical model that delivers useful inference (see also section 6). Again we first analyze M_1^R as this captures the essentials and then briefly a generalized version of M_2^R .

From (23) and the corresponding root, it is clear that a unit root occurs when $a_1 + a_3 = 0$ and/or $c_1 = 0$. Equivalently, existence of the steady state in (6) is lost. In terms of Figure 1, $a_1 + a_3 = 0$ means that the curves in the left panel are parallel and hence (generically) there is no equilibrium real wage where $b = d$. Population becomes independent of income as the effect on b is the same as on d . We find that,

$$\Pi = \alpha\beta', \quad \alpha = \begin{pmatrix} -c_1 & 0 \\ a_3c_1 & -1 \\ 1 + a_3c_1 & -1 \end{pmatrix}, \quad \beta = \begin{pmatrix} 0 & a_3 \\ 1 & 1 \\ -1 & 0 \end{pmatrix}, \quad (29)$$

i.e. there are two identified cointegrating vectors. The first implies a stationary growth rate of population, and the second is a "check" relation.

In terms of M_1^S , $\ln N_t$ becomes a random walk with a drift $\mu_0 = a_0 - a_2$, while $\ln A_t$ is still the same random walk (see 18). Thus, the processes are independent I(1) processes, and not cointegrating anymore as $\Pi = 0$ in (19).

If $c_1 = 0$ the labor demand schedule is horizontal and income becomes independent of population. In such a case no equilibrium for w , and therefore for b and d , exists since shocks w are not "corrected" as the growing population has no effect on wages. We find that,

$$\Pi = \alpha\beta', \quad \alpha = \begin{pmatrix} 0 & 0 \\ -1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \beta = \begin{pmatrix} -a_1 & a_3 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}. \quad (30)$$

In this case the two Malthusian check relations become identified cointegrating vectors. The α matrix in (30) implies that the real wage becomes weakly exogenous, which in this case can be interpreted as exogenous in the usual economic sense (See e.g. Møller 2008). It implies that the real wage is the common stochastic trend, pushing the long-run paths of the rates b_t and d_t . These adjust passively, and in turn, have no effect on the real wage - since the marginal

productivity effect (c_1) of population on income is zero.

Now, M_1^S changes radically. First the normalization in (18), $\frac{1}{c_1}$ is invalid. Instead, we find that $w_t = c_0 + \ln A_t$, i.e. a constant plus a random walk and hence a random walk. This means that the rates b_t and d_t also become random walks, that do not cointegrate 1:1 (homogeneously), hence the population growth rate $b_t - d_t$ is I(1), so that by definition population becomes I(2). This is seen explicitly from the MA-representation,

$$\ln N_t = \ln N_0 + b_0 t + (a_1 + a_3) \sum_{i=1}^t \sum_{j=1}^{i-1} \varepsilon_{At} + \sum_{i=1}^t \varepsilon_{Ni}, \quad (31)$$

where $b_0 \equiv (a_0 - a_2) + (a_1 + a_3)(c_0 + \ln A_0)$. Hence the shocks $\varepsilon_{bt} - \varepsilon_{dt}$ generate an I(1) stochastic trend, while the labor demand shocks generate a second order or I(2) trend. In addition to these linear stochastic trends there will also be a linear deterministic trend, which will be present even if $a_0 = a_2$ simply due to the initial value $\ln A_0$. Note that population is I(2) but labor demand is still only I(1).

The general model, \tilde{M}_2^R , in Appendix A.2 which is relevant for the estimation, has,

$$\alpha = \begin{pmatrix} 0 & 0 & -c_1 \\ -f_0(1-\delta) & -(1-\gamma-\phi) & -\phi - \eta_0 c_1 f_0 \\ -\tilde{f}(1-\lambda) & -(1-\lambda) & 1 - \lambda - \psi_0 c_1 \\ -(1-\delta) & 0 & -\eta_0 c_1 \end{pmatrix}, \quad \beta' = \begin{pmatrix} -\tilde{a}_5 & 0 & 0 & 1 \\ 0 & 1 & 0 & -\tilde{f} \\ 0 & 1 & -1 & 0 \end{pmatrix}, \quad (32)$$

in the case when $\tilde{a}_1 + \tilde{a}_3 = 0$, and when $c_1 = 0$ we get,

$$\alpha = \begin{pmatrix} 0 & 0 & 0 \\ \frac{(\phi\tilde{a}_3 - \eta f_0)}{\tilde{a}_5} & -(1-\gamma) & \phi \\ 0 & 0 & -(1-\lambda) \\ -(1-\delta) & 0 & 0 \end{pmatrix}, \quad \beta' = \begin{pmatrix} -\tilde{a}_5 & 0 & 0 & 1 \\ 0 & 1 & 0 & \frac{\phi\tilde{a}_3}{(1-\gamma)\tilde{a}_5} - \frac{(1-\gamma-\phi)}{(1-\gamma)}\tilde{f} \\ \tilde{a}_3 & 0 & 1 & 0 \end{pmatrix}. \quad (33)$$

In both (32) and (33), the reduced rank, r , is exactly three as suggested by the data, and β is identified.

4.2 Imposing the persistence restrictions

Under the restriction $\tilde{a}_1 + \tilde{a}_3 = 0$ we leave α unrestricted, and hence only impose the restrictions on β . Under $c_1 = 0$ we also impose the zero row in α in addition to the β restrictions. Conditional on the estimated β , one can then impose the additional restrictions on α (and Γ_i, Φ) and conduct Gaussian based inference. We do not pursue this here.

Our preferred specification is S_3 with $k = 3$. Data suggested that the constant is restricted to the span of α , implying insignificant deterministic trends as expected from Figure 2.

Table 2: Testing the first assumption of Malthusian persistence

α				β'								
Δw_t	0.03 [3.77]	0.01 [1.31]	-0.00 [-1.79]	w_t	b_t	d_t	m_t	D_t^P	Ds_t^{1643}	Ds_t^{1659}	$Const.$	
Δb_t	-0.20 [-1.88]	-0.37 [-5.25]	-0.02 [-0.77]	β'_1	-9.16 [-9.11]	0.00	0.00	1.00	6.15 [3.96]	0.92 [1.95]	2.26 [4.56]	7.14 [3.96]
Δd_t	0.37 [1.51]	-0.28 [-1.69]	0.34 [5.30]	β'_2	0.00	1.00	0.00	-1.74 [-9.12]	-8.86 [-3.27]	2.87 [3.58]	-3.96 [-5.49]	-16.86 [-9.50]
Δm_t	-0.27 [-4.65]	-0.04 [-0.93]	-0.05 [-3.41]	β'_3	0.00	1.00	-1.00	0.00	6.26 [0.89]	5.06 [4.62]	0.00	-7.96 [-9.36]

The estimates under the restriction $\tilde{a}_1 + \tilde{a}_3 = 0$ are given in Table 2 (t -values in brackets). The restriction was accepted with a p -value of 65%. The estimates of \tilde{a}_5 and \tilde{f} are respectively, 9.16 and 1.74 (see 32), and both strongly significant, completely in line with theory. The α matrix shows the desired significant error-correction in m to the first relation (-0.27), and in b to the second relation (-0.37), while adjustment in b to the third relation is insignificant and happens through d instead (0.34). The estimates α_{12} and α_{42} are insignificant as expected⁸. The estimate of α_{32} suggests error correction but it is borderline insignificant (t -value = -1.69). The estimate of α_{43} (-0.05) is also consistent with α in (32).

There are however, some surprises. First, the estimate of α_{13} , i.e. of c_1 , is (borderline) insignificant. It is possible that it may become significant when the model is cleaned from superfluous regressors. However, what is interesting is that the estimate is rather small, 0.004 ("0.00" in output). This suggests that $c_1 \approx 0$, i.e. that the other reason for persistence could also be relevant. Second, the estimate of α_{11} suggests significant error correction of real wages to the first relation which is not predicted from theory, and does not seem readily interpretable.

In spite of these "surprises" the estimated α and β seem rather clear-cut when compared to the matrices in (32). However, the restriction $\tilde{a}_1 + \tilde{a}_3 = \tilde{a}_5 \tilde{f} + \tilde{a}_3 = 0$, may seem hard to accept, since it implies that the long-run effect from income on the death rate, $-\tilde{a}_3$, is positive, when $\tilde{a}_5 \tilde{f} > 0$, which seems plausible and indeed empirically supported. So, unless we have convincing theories that suggest just the opposite of positive checks, it is not clear that we should accept these results at face value. It should be intuitively clear that even if the positive check coefficient has the right sign but is very small, then unless the preventive check mechanism is sufficiently strong, we are likely to get away with forcing this restriction upon the model.

Finally, the results in Table 2 as well as an individual test for stationarity of the population growth rate, $b_t - d_t$ (corrected for shift) with a p -value of 63%, support the use of the linear approximation cf. the discussion in section 2.2.

Imposing the restriction that $c_1 = 0$, we get the estimates in Table 3. This restriction was also accepted with a p -value at 10%. Recalling (33) we see that the first relation (preventive check relation) has the right sign on \tilde{a}_5 and it is significant, and m is error-correcting with high significance (t -value is -6.89).

⁸ α_{ij} in i 'th row, j 'th column of α .

Table 3: Testing the second assumption of Malthusian persistence

α				β'								
Δw_t	0.00	0.00	0.00	w_t	b_t	d_t	m_t	D_t^P	Ds_t^{1643}	Ds_t^{1659}	<i>Const.</i>	
Δb_t	-0.28 [-2.15]	-0.37 [-5.92]	0.03 [1.22]	β'_1	-5.41 [-5.92]	0.00	0.00	1.00	5.30 [3.58]	0.00	2.20 [7.71]	0.53 [0.32]
Δd_t	0.56 [1.87]	-0.04 [-0.26]	-0.42 [-6.35]	β'_2	0.00	1.00	0.00	-2.05 [-9.21]	-10.29 [-3.45]	3.68 [4.27]	-5.04 [-6.10]	-14.05 [-6.78]
Δm_t	-0.49 [-6.89]	-0.11 [-3.05]	0.04 [2.25]	β'_3	0.90 [0.23]	0.00	1.00	0.00	-7.85 [-1.20]	-4.75 [-3.93]	0.00	-26.23 [-3.68]

The estimate of ϕ (α_{23}) is insignificant, and hence $\phi = 0$ cannot be rejected. From β in (33) we see that this restriction identifies the last element of the second cointegrating vector as \tilde{f} . The estimate is -2.05 and significant, and hence, very much in line with the -1.74 in Table 2. Here, the error-correction to this relation is in b as it should and also very significant. The third relation - the positive check relation - now shows us that the positive check may have occurred, but it was too weak to be significant (t -value is 0.23). Finally, as expected, the death rate is indeed error-correcting to this relation. When restricting \tilde{a}_3 to zero the p -value rose to 17% , while the estimates and t -values were basically unaltered.

Both estimated matrices α and β are again very clear-cut. The bulk of significance of the adjustment coefficients takes place in the variables on which we have normalized (the most significant coefficients are α_{41} , α_{22} and α_{33}). In addition the rest of the α coefficients are in fact interpretable: First, the estimate of α_{42} is significantly negative which is probably capturing the purely demographic effect, that when b_{t-1} rises, N_t goes up, which given M_t , means that $m_t = \frac{M_t}{N_t}$ will fall. Likewise, the positive estimate of α_{43} is also the automatic effect, that when d_{t-1} rises then $m_t = \frac{M_t}{N_t}$ rises simply because N_t always falls more than M_t .

To sum up, the results are remarkably clear-cut, and under both restrictions data suggest that the positive check, \tilde{a}_3 , is insignificant, that \tilde{f} , the long-run effect from marriages on births, is around 2 and highly significant, and that \tilde{a}_5 is probably around 5-9 and also highly significant. Error-correction is also as expected and very significant. The estimate in Table 2, $\hat{c}_1 = 0.004$, and the strikingly clear picture in Table 3 suggest that c_1 is negligible. In terms of Figure 1, the data thus support a horizontal death rate schedule which fluctuates in a stationary manner around a level that shifts under extraordinary circumstances. The birth rate relation is upward sloping, while the labor demand schedule is horizontal and shifts like a random walk, determining the position on the birth rate schedule. Finally, $c_1 \approx 0$ means that population, and hence the labor supply schedule changes even more persistently, i.e. like I(2).

It is interesting to note that in the analysis of Lee and Anderson (2002), the error term in the intercepts for the preventive and positive check schedules is assumed to be I(1). It can be seen clearly from our equation 2 that if ε_{bt} were an I(1)-process, then it would be impossible for the I(1) processes b_t and w_t to cointegrate. This is however what our results convincingly support, as the m_t cointegrates with w_t and b_t cointegrates with m_t .

Even though both restrictions of Malthusian persistence are accepted we choose to focus on the latter ($c_1 = 0$). First of all, as argued above, the restriction $\tilde{a}_1 + \tilde{a}_3 = 0$ may seem hard to

accept per se. Secondly, under this restriction we saw that the results pointed towards $c_1 = 0$ with the almost significant estimate $\hat{c}_1 = 0.004$. Thirdly, the strongly significant preventive checks working through marriages combined with insignificant positive checks have been found previously in Lin Lee and Loschky (1987). Finally, the estimate of α in Table 3 was the most interpretable. However, as lack of diminishing marginal returns, $c_1 \approx 0$, may also seem difficult to accept, we now consider a broader interpretation of this parameter.

5 A tentative interpretation of the empirical results

The previous analysis clearly suggests that the marginal productivity effect, c_1 , is very small. Under the Cobb-Douglas technology, c_1 is the sum of the income shares to capital and land, so unless we are willing to accept an estimate of this at 0.004, from Table 2, we must adopt a broader interpretation of c_1 . By introducing only one more parameter we show how $\ln A$, can be endogenized to produce a positive effect from population on real wages via "increased technology". This counteracts the negative marginal productivity effect, and results in a "modified c_1 " - broadly interpreted as the net effect of the two forces.

5.1 A Malthusian model with endogenous technological progress

That larger populations might give rise to more technological progress was suggested long ago by Adam Smith. He suggested that larger populations – in his words a greater “extent of the market” - allow for greater division of labor and thus increased opportunities for learning-by-doing (see the discussion in Persson 1988). More recently, Boserup (1965) suggested the possibility that larger population density might put such a strain on resources that new technologies are brought into use in order to maintain a sufficient food supply.

There are many ways to model feed-back effects from population on technology, and possibly many ways to interpret Smith and/or Boserup. Here, we suggest a simple and tentative way to incorporate such mechanisms into the framework above, while we leave the economic interpretation relatively open. We suggest a slight generalization to allow for "Boserupian/Smithian" effects, implying that the pure Malthusian model becomes a special case of this model. We refer to our model as the *Synthetic model*, and use the understood notation, \mathcal{M}_1 , \mathcal{M}_1^R and \mathcal{M}_1^S .

It suffices to consider the extension of M_1 . The Synthetic model, \mathcal{M}_1 , is simply M_1 with (4) substituted by the two equations,

$$\ln A_t = c_3 \ln N_t + \ln X_t, \tag{34}$$

$$\ln X_t = \ln X_{t-1} + \varepsilon_{Xt}, \tag{35}$$

where $c_3 \geq 0$, X_t could be general or basic knowledge and A_t is knowledge that matters directly for production. We note that \mathcal{M}_1 collapses to M_1 when $c_3 = 0$. In this case there is no distinction between A_t and X_t , and we are back in the Malthusian case. The Boserupian/Smithian

mechanism arises when $c_3 > 0$, in which case there will be a positive effect from the level of population on the level of applied technology.

As previously, to learn about dynamics we derive \mathcal{M}_1^R and \mathcal{M}_1^S . First we get for \mathcal{M}_1^R ,

$$\mathcal{M}_1^R : \Pi = \begin{pmatrix} 0 & -\bar{c}_1 & \bar{c}_1 \\ a_1 & -(1 + a_1\bar{c}_1) & a_1\bar{c}_1 \\ -a_3 & a_3\bar{c}_1 & -(1 + a_3\bar{c}_1) \end{pmatrix}, \quad \det(\Pi) = \bar{\alpha}_1, \quad (36)$$

where $\bar{\alpha}_1 = -\bar{c}_1(a_1 + a_3)$, and the "broadly interpreted" c_1 is,

$$\bar{c}_1 \equiv c_1 - c_3, \quad (37)$$

resembling \mathcal{M}_1^R completely. Hence, we can retain the core classical assumption of diminishing marginal returns to labor, $c_1 > 0$, and still have a small, and hence insignificant \bar{c}_1 consistent with the evidence. This approach is further discussed in section 6.

Concerning \mathcal{M}_1^S we get the generalized counterpart to (18),

$$\begin{aligned} \Delta \ln N_t &= \mu_0 + \alpha_1(\ln N_{t-1} - \beta_1 \ln A_{t-1}) + \varepsilon_{Nt}, \\ \Delta \ln A_t &= \bar{\mu}_0 + \alpha_2(\ln N_{t-1} - \beta_1 \ln A_{t-1}) + \bar{\varepsilon}_{At}, \end{aligned} \quad (38)$$

where $\alpha_1 = -c_1(a_1 + a_3)$, $\alpha_2 \equiv c_3\alpha_1$, $\bar{\mu}_0 \equiv c_3\mu_0$ and $\bar{\varepsilon}_{At} \equiv c_3\varepsilon_{Nt} + \varepsilon_{Xt}$.

The crucial difference between (38) and (18) is that technology is no longer weakly exogenous but is indeed adjusting to population in the long run, with the coefficient α_2 . So, there is still the same long run equilibrium relationship between technology and population, $\beta'x_t \equiv \ln N_t - \beta_1 \ln A_t$. However, the random walk path that both variables follow (so that their "difference", $\beta'x_t$, becomes stationary) is no longer determined solely by the cumulation of $\ln A$ shocks. The common trend is now $\alpha'_\perp \sum_{i=1}^t \varepsilon_i = \sum_{i=1}^t \varepsilon_{Ni} - \frac{1}{c_3} \sum_{i=1}^t \bar{\varepsilon}_{Ai}$, i.e. it includes *both* technology shocks, $\bar{\varepsilon}_A$ and population shocks, ε_N , as is intuitively clear since both variables have a long-run influence on each other.

Given the Malthusian core assumptions and that $c_3 > 0$, the condition that the rates are jointly stationary, and the stock variables I(1) and cointegrating with $\beta' = (1, -\beta_1)$, is naturally that $c_3 < c_1$. That is, the Boserupian/Smithian effect must not dominate the Malthusian.

There is an important difference to note between our Synthetic model and consensus literature. For example, within the interpretations of Boserup and the literature of (unified) growth theory there has been an emphasis on explaining the *growth* of technology with the *level* of the population (See Galor 2005). This is not what our Synthetic model says. Here, the dependence is between the *levels* ($\ln N_t$ and $\ln A_t$). But admittedly, more analysis is needed, which combines the flexible cointegrated VAR with advanced growth theory, for instance along the lines of Aiyar, Dalgaard, and Moav (2008).

Given this interpretation, our empirical results can be viewed as the limit case of the Synthetic model when $\bar{c}_1 = 0$ (or small). The evidence is thus consistent with a Malthusian process (with preventive checks but no positive checks), which is counteracted by Boserupian

- and/or Smithian effects from population size on technology, to the extent that the net-effect of population on income becomes statistically insignificant.

6 Discussion

The result that c_1 , or rather \bar{c}_1 , is zero implies that $b_t - d_t$ and m_t are $I(1)$, and hence non-stationary, which makes the linearization in (14) problematic. However, $\bar{c}_1 = 0$ should probably be interpreted as an approximation of \bar{c}_1 close to zero. For a persistent process, such an approximation may deliver a statistical model which facilitates inference (see section 2.3.1 and Johansen 2006). In this sense, the unit root - (cointegration) analysis in section 4.1-5 is consistent with an underlying stationary but very persistent process, and hence the linearization can be justified.

As mentioned, our interpretation in section 5 should be regarded as tentative. In particular, the problem with the Synthetic model is that it does not identify c_1 and c_3 only their difference \bar{c}_1 . This implies that *based on the data*, we cannot conclude whether both effects, c_1 and c_3 , do in fact operate but simply offset each other - as we assume above, or whether they are both negligible. In spite of this and its ad hoc nature, we believe that our synthetic model remains interesting, as it shows how to incorporate mechanisms advocated by the traditional adversaries of the Malthusian ideas. It is done in a simple fashion and retains the core economic assumption of diminishing returns to labor. Moreover, it may serve as a point of departure for developing models that do identify c_1 and c_3 .

There are a number of additional interesting implications from our empirical results. First, the acceptance of w being weakly exogenous together with the first rows of the unrestricted estimates of $\hat{\Gamma}_1$ and $\hat{\Gamma}_2$ (Appendix C), suggests that w can be approximated empirically as a random walk. Hence, it can be regarded as exogenous in the usual economic sense (See e.g. Møller 2008).

We also note that if the unobserved $\ln A_t$ were better described by a more general $I(1)$ process, i.e. an $AR(\kappa)$ process with $z = 1$, this would imply MA-errors in the real wage equation and hence, autocorrelated residuals. As we have seen, this is however rejected convincingly. Such $AR(\kappa)$ unit root processes could for example capture overshooting or gradual increase in $\ln A$, both potentially relevant descriptions of technical knowledge.

The rejection of autocorrelation also implies that changes in $\ln A$ impact rather rapidly on labor demand and real wages. This is so since, if $\ln A$ had a more gradual impact, say due to gradual diffusion of knowledge, this would entail distributed lags of the unobserved $\ln A$ in the real wage equation, that is, MA-errors, and thus autocorrelated residuals.

As is natural to expect, births may affect labor supply, and hence wages, with a rather long lag, say, 15-25 years, as there must be a lower limit for the age at labor force entrance and employment. For example, Nicolini sought to handle this problem by the inclusion of population lagged 25 years (Nicolini 2007). What is interesting is that our results do *not* suggest this, since if such long lags were operating, the unrestricted estimates of the real wage equation would

probably suggest this: The included first lags of b , and the lagged differences, Δb_{t-i} , would be significant. In addition, autocorrelated real wage residuals would occur if we have fitted too few lags. From the estimate of α , the insignificant first rows of $\widehat{\Gamma}_i$ in Appendix C, and the autocorrelation tests in Appendix B, it is clear that none of this is found in our empirical results. A possible, and to our minds overlooked explanation, is that births may in fact impact immediately on labor supply (and hence real wages), since having a child implies yet another mouth to feed, which may induce the breadwinner to work more. When the child eventually becomes old enough, it enters employment and so to speak, takes over the effect on aggregate labor supply (measured in hours). Regardless of what happens under transition the long-run effect, which is our concern, is the same - labor supply will increase, and real wages fall, given the initial rise in population.

The estimated real wage equation has further interesting implications: First, the previous analysis ignores the role of net-migration. One way to include this could be to generalize the population equation, (5), to,

$$\ln N_t = \ln N_{t-1} + b_{t-1} - d_{t-1} + \phi_N D_{Nt} + e_{Nt}, \quad (39)$$

where $\phi_N D_{Nt} + e_{Nt}$ describes net-migration. The term, D_{Nt} , is an impulse dummy, capturing large changes in net-migration, whereas e_{Nt} is a residual term. Hence, compared to Lee and Anderson (2002) we have added D_{Nt} . Furthermore, the equation for labor demand, (4), could be generalized to,

$$\ln A_t = c_A + \ln A_{t-1} + \phi_A D_{At} + \varepsilon_{At}, \quad (40)$$

where c_A captures regular growth of technology and/or capital per worker, for example. The term, $\phi_A D_{At}$, again with D_{At} being an impulse dummy, is supposed to capture the effect from major technological improvements if such were to occur.

Now, these equations imply the real wage equation,

$$\Delta w_t = c_A - c_1(b_{t-1} - d_{t-1}) - c_1\phi_N D_{Nt} + \phi_A D_{At} + \varepsilon_t^w, \quad (41)$$

where $\varepsilon_t^w \equiv \varepsilon_{At} - c_1 e_{Nt}$. From (41), we see that the empirical results (that w_t can be approximated with a random walk) suggest: 1) No technological smooth/deterministic growth. 2) No major technological improvements took place during the period, since no dummies were needed in the Δw equation. Furthermore, since c_1 (or \bar{c}_1) is zero this implies the identification of the ε_{At} from the wage shock, ε_t^w , and that net-migration does not matter for the analysis. The latter is quite fortunate since reliable data on net-migration do not exist (see the discussion in Bailey and Chambers 1993).

7 Constancy of the parameters and the ‘Malthusian era’

Our analysis cuts the sample off at 1760, and thus excludes the period often associated with the Industrial Revolution, at which point it might be expected that the Malthusian mechanisms

break down. We now want to assess the assumption of constant parameters within the period 1560-1760. If this seems reasonable, we may furthermore estimate the model recursively for the samples, 1560- T , $T = 1761, 1762, \dots$, to determine the period in which our results in Table 3 hold.

To assess the assumption of constant parameters within our sample, a baseline sample is chosen, 1560- T^B . The model is then recursively estimated for the samples, 1560- $T^B + 1$, 1560- $T^B + 2, \dots$ 1560-1760, and the evolution of the estimates is mapped out (Dennis, Hansen, and Juselius 2006).

First, α and β are recursively estimated. The relevant α coefficients are significantly error correcting and of a similar level for all sample lengths. The β coefficients are likewise remarkably constant. Hence, the mechanisms we have identified are stable and operating at least until the industrial revolution.

Extending the sample beyond 1760, the recursively estimated α and β coefficients seem stable until about 1785. At this point, d_t ceases to error-correct to the third relation. The marriage rate in the first relation and the birth rate in the second relation, on the other hand, continue to error-correct for almost all sample lengths, although the coefficients associated with both fall. The β coefficient on w_t in the first relation also remains very constant as does that on the death rate in the second relation. However, the coefficient on m_t becomes insignificant from around about 1800.

These results imply that industrialization does impact on the Malthusian relationships, as expected. The weak (insignificant) positive check disappears soon after the onset of industrialization. However, the preventive check seems to persist longer. Indeed, the estimate of \tilde{a}_5 appears remarkably robust and significant throughout the industrialization period (as hypothesized by Sharp and Weisdorf 2008). The automatic effect of the death rate on the birth rate is also very robust, as might be expected, but m ceases to have an impact on b after 1800, thus destroying the preventive check mechanism. The positive check – if ever present – disappeared much earlier.

Our results differentiate themselves in several important respects from those found in previous studies. As we do not attempt to estimate a model for the period both before and after industrialization, we obtain a statistically well-defined - and remarkably stable model. By including m , we are able to demonstrate that the dependence of the marriages on real wages continues beyond industrialization. In contrast to Nicolini (2007), we find evidence that the workings of the preventive check remained remarkably constant until about 1800. Our results confirm the findings of Bailey and Chambers (1993), who also find a positive effect of the real wage on fertility and nuptiality, and a negative impact of fertility from the death rate for the years until 1800.

8 Conclusions

Our empirical results are consistent with a Malthusian process - with preventive checks working through marriages but negligible positive checks - which is counteracted by Boserupian and/or Smithian effects from population size on technology, to an extent that the net-effect of population on income becomes statistically insignificant. This supports Malthus' modified views in the second edition of his Essay from 1803. He named a number of factors which would offset diminishing returns to labor, including increased investment and improvements in agriculture (see e.g. Collard 2001).

As a by-product of our analysis we do not find any support for Malthusian Oscillations, as opposed to Lin Lee and Loschky (1987). Furthermore, the evidence is consistent with no linear deterministic trend in technology, and allows technological regress (see e.g. Aiyar, Dalgaard, and Moav 2008). Finally, recursive estimation suggests that the identified model operates until 1785, and that the preventive check mechanism (via marriages) persists until about 1800.

A number of promising avenues for future research open up in the wake of this paper. First, as discussed in section 6 it should be both possible and useful to develop the Synthetic model, or similar models, in order to identify both the Malthusian effect, c_1 , and the counteracting Boserupian effect, c_3 , and not just their difference, \bar{c}_1 . Second, due to our professional background, it is probably possible to improve the demographic foundation of our model. Third, there is scope for economic theories explaining why the real wage is almost completely exogenous. Fourth, given data availability, it should be possible to apply our approach to other pre-industrial economies, and/or potentially current developing countries.

Appendix

8.1 Appendix A

8.1.1 A.1. Analysis of M_2

M_2^s with $s > 1$.

The model, M_2 , is,

$$w_t = c_0 - c_1 \ln N_t + \ln A_t, \quad (42)$$

$$\begin{aligned} b_t \simeq & e_0 + e_1 b_{t-1} - e_1 d_{t-1} + f_1 m_{t-1} + e_2 b_{t-2} - e_2 d_{t-2} + f_2 m_{t-2} + \dots \\ & \dots + e_s b_{t-s} - e_s d_{t-s} + f_s m_{t-s} + \varepsilon_{bt}, \end{aligned} \quad (43)$$

$$d_t = a_2 - a_3 w_t + \varepsilon_{dt}, \quad (44)$$

$$m_t = a_4 + a_5 w_t + \varepsilon_{mt}, \quad (45)$$

$$\ln A_t = \ln A_{t-1} + \varepsilon_{At}, \quad (46)$$

$$\ln N_t = \ln N_{t-1} + b_{t-1} - d_{t-1}. \quad (47)$$

As when deriving (21) we first derive the generalized birth relation by inserting (45) in (43) to get,

$$b_t = a_0^* + e_1 \Delta \ln N_t + \dots + e_s \Delta \ln N_{t-(s-1)} + a_{11}^* w_{t-1} + a_{12}^* w_{t-2} + \dots + a_{1s}^* w_{t-s} + \varepsilon_{bt}^*, \quad (48)$$

where $a_0^* \equiv e_0 + a_4 f$, $f \equiv f_1 + f_2 + \dots + f_s$, $a_{1j}^* \equiv f_j a_5 > 0$, $j = 1, \dots, s$ and $\varepsilon_t^* \equiv f_1 \varepsilon_{mt-1} + \dots + f_s \varepsilon_{mt-s} + \varepsilon_{bt}$. Hence, we see that MA-errors occur. This of course does not occur in models where m is included, in particular M_2^R and the full system. Proceeding as in the case $s = 1$, we arrive at the following CVAR,

$$\Delta \ln N_t = \mu_0^* + \alpha_1^* (\ln N_{t-1} - \beta_1 \ln A_{t-1}) + \theta_{11}^* \Delta \ln N_{t-1} + \dots + \theta_{1s}^* \Delta \ln N_{t-s}, \quad (49)$$

$$+ \theta_{21}^* \Delta \ln A_{t-1} + \dots + \theta_{2s}^* \Delta \ln A_{t-s} + \varepsilon_{Nt}^*, \quad (50)$$

$$\Delta \ln A_t = \varepsilon_{At},$$

where,

$$\mu_0^* \equiv (a_0^* - a_2) + (a_1^* + a_3)c_0, \quad \alpha_1^* \equiv -c_1(a_1^* + a_3), \quad a_1^* \equiv a_{11}^* + \dots + a_{1s}^* > 0, \quad (51)$$

$$\theta_{1j}^* \equiv e_j + c_1 \sum_{l=j}^s a_{1l}^*, \quad \theta_{2j}^* \equiv -\sum_{l=j}^s a_{1l}^*, \quad j = 1, \dots, s \text{ and } \varepsilon_{Nt}^* \equiv \varepsilon_{t-1}^* - \varepsilon_{dt-1},$$

As we can see the long-run dynamics has exactly the same structure as in the simple models M_1^S . The cointegrating vector is identical, the adjustment coefficient, α_1^* , is negative because $c_1 > 0$, $a_5 > 0$, $a_3 > 0$ and all $f_j > 0$, i.e. due to the three crucial Malthusian assumptions. As with M_2^S , the transitory or adjustment dynamics have just become more complicated, otherwise the model is basically the same. Note that $\det(\alpha'_\perp \Gamma \beta_\perp) \neq 0$ always.

M_2^R with $s > 1$.

In the usual ECM form we get,

$$\Pi = \begin{pmatrix} 0 & -c_1 & c_1 & 0 \\ 0 & -(1 - \bar{e}_1) & -\bar{e}_1 & f \\ -a_3 & a_3 c_1 & -a_3 c_1 - 1 & 0 \\ a_5 & -a_5 c_1 & a_5 c_1 & -1 \end{pmatrix}, \quad \det(\Pi) = -\alpha_1^*, \quad (52)$$

where $\bar{e}_1 \equiv \sum_{i=1}^s e_i$ and $f = \sum_{j=1}^s f_j$ (defined above).

The model can be written as,

$$\Delta x_t = \Pi x_{t-1} + D_1 \Delta x_{t-1} + D_2 \Delta x_{t-2} \dots + D_{s-1} \Delta x_{t-(s-1)} + \mu + v_t, \quad (53)$$

with Π defined in (52), and D_i ,

$$D_i \equiv \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & -\bar{e}_{i+1} & \bar{e}_{i+1} & \frac{e_{i+1}}{m^*} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad i = 1, \dots, (s-1), \quad (54)$$

with $\bar{e}_j \equiv \sum_{l=j}^s e_l$, $j = 2, \dots, s$. For later we note that,

$$\Gamma \equiv \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 + \sum_{j=2}^s \bar{e}_j & -\sum_{j=2}^s \bar{e}_j & -\frac{1}{m^*} \bar{e}_2 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad (55)$$

where we use the matrix notation from Johansen (1996).

8.1.2 A.2. The generalized version of M_2^R , \widetilde{M}_2^R

Generalizations relative to M_2^R : The wage equation: none. The birth rate equation: a current effect from m_t (f_0) and other effects from lagged b and d , γ_i and ϕ_i , in addition to the e_i coefficients in (14). The death rate equation: More gradual adjustment by the ψ_i - and λ_i coefficients. The marriage rate equation: More gradual adjustment by the η_i - and δ_i coefficients.

The equations are,

$$w_t = w_{t-1} - c_1(b_{t-1} - d_{t-1}) + \varepsilon_{At}, \quad (56)$$

$$b_t = e_0 + f_0 m_t + \gamma_1 b_{t-1} + \phi_1 d_{t-1} + f_1 m_{t-1} + \dots + \gamma_k b_{t-k} + \phi_k d_{t-k} + f_k m_{t-k} + \varepsilon_{bt}, \quad (57)$$

$$d_t = a_2 + \psi_0 w_t + \lambda_1 d_{t-1} + \dots + \psi_k w_{t-k} + \lambda_k d_{t-k} + \varepsilon_{dt}, \quad (58)$$

$$m_t = a_4 + \eta_0 w_t + \delta_1 m_{t-1} + \dots + \eta_k w_{t-k} + \delta_k m_{t-k} + \varepsilon_{mt}. \quad (59)$$

This implies the Π matrix (see 27),

$$\Pi = \begin{pmatrix} 0 & -c_1 & c_1 & 0 \\ f_0 \eta & -(1-\gamma) - \eta_0 c_1 f_0 & \phi + \eta_0 c_1 f_0 & f_0 \delta + \bar{f} \\ \psi & -\psi_0 c_1 & -(1-\lambda) + \psi_0 c_1 & 0 \\ \eta & -\eta_0 c_1 & \eta_0 c_1 & -(1-\delta) \end{pmatrix}, \quad (60)$$

where $\bar{f} = \sum_{i=1}^k f_i$, $\eta \equiv \sum_{i=0}^k \eta_i$, $\gamma \equiv \sum_{i=1}^k \gamma_i$, $\phi \equiv \sum_{i=1}^k \phi_i$, $\delta \equiv \sum_{i=1}^k \delta_i$, $\lambda \equiv \sum_{i=1}^k \lambda_i$, $\psi \equiv \sum_{i=0}^k \psi_i$. The determinant of this can be written as,

$$\det(\Pi) = c_1 \left(\tilde{a}_3 + \tilde{a}_5 \bar{f} \right) (1-\lambda)(1-\delta)(1-\gamma-\phi), \quad (61)$$

provided that $(1 - \lambda) \neq 0, (1 - \delta) \neq 0, (1 - \gamma - \phi) \neq 0$, and where,

$$\tilde{f} \equiv \frac{\sum_{i=0}^k f_i}{(1 - \gamma - \phi)}, \tilde{a}_3 \equiv \frac{-\psi}{1 - \lambda}, \tilde{a}_5 \equiv \frac{\eta}{(1 - \delta)}. \quad (62)$$

The special case M_2^R is obtained by setting $\gamma_i = e_i, \phi_i = -\gamma_i, f_0 = 0$ and $k = s$ in (57). In (58) we set $\psi_0 = -a_3, \psi_i = 0$ and $\lambda_i = 0$ for $i = 1, 2..k$. In (59) we set $\eta_0 = a_5, \eta_i = 0$ and $\delta_i = 0$ for $i = 1, 2..k$.

Appendix B

B.1. Misspecification tests. Model: S_1

Tests for Autocorrelation

LM(1):	ChiSqr(16)	=	16.74	[0.40]
LM(2):	ChiSqr(16)	=	12.17	[0.73]
LM(3):	ChiSqr(16)	=	16.97	[0.39]
LM(4):	ChiSqr(16)	=	11.33	[0.79]

Test for Normality:	ChiSqr(8)	=	81.91	[0.00]
---------------------	-----------	---	-------	--------

Test for ARCH:

LM(1):	ChiSqr(100)	=	230.86	[0.00]
LM(2):	ChiSqr(200)	=	337.67	[0.00]
LM(3):	ChiSqr(300)	=	447.58	[0.00]
LM(4):	ChiSqr(400)	=	569.63	[0.00]

Univariate Statistics

	Mean	Std.Dev	Skewness	Kurtosis	Maximum	Minimum
DLYBL	0.00	0.09	-0.06	3.72	0.31	-0.28
DCBR	0.00	1.16	-0.24	3.53	2.99	-3.70
DCDR	0.00	3.23	1.68	8.76	17.49	-6.93
DCMR	-0.00	0.68	0.39	4.18	2.33	-2.03

	ARCH(4)	Normality	R-Squared
DLYBL	6.79 [0.15]	5.71 [0.06]	0.22
DCBR	4.12 [0.39]	4.08 [0.13]	0.67
DCDR	2.97 [0.56]	60.92 [0.00]	0.37
DCMR	13.20 [0.01]	10.86 [0.00]	0.34

B.2. Misspecification tests. Model: S_2

RESIDUAL ANALYSIS

Tests for Autocorrelation

LM(1): ChiSqr(16) = 15.40 [0.50]
LM(2): ChiSqr(16) = 7.54 [0.96]
LM(3): ChiSqr(16) = 11.86 [0.75]
LM(4): ChiSqr(16) = 14.92 [0.53]

Test for Normality: ChiSqr(8) = 36.22 [0.00]

Test for ARCH:

LM(1): ChiSqr(100) = 202.89 [0.00]
LM(2): ChiSqr(200) = 341.02 [0.00]
LM(3): ChiSqr(300) = 458.50 [0.00]
LM(4): ChiSqr(400) = 582.85 [0.00]

Univariate Statistics

	Mean	Std.Dev	Skewness	Kurtosis	Maximum	Minimum
DLYBL	0.00	0.09	-0.04	3.71	0.31	-0.27
DCBR	0.00	1.14	-0.16	3.54	3.25	-3.55
DCDR	0.00	2.59	0.71	4.10	8.30	-6.71
DCMR	-0.00	0.68	0.37	3.98	2.22	-2.01

	ARCH(4)		Normality		R-Squared
DLYBL	7.19	[0.13]	5.61	[0.06]	0.23
DCBR	2.24	[0.69]	3.90	[0.14]	0.68
DCDR	19.04	[0.00]	14.66	[0.00]	0.59
DCMR	15.71	[0.00]	8.70	[0.01]	0.36

B.2. Misspecification tests. Model: S_3

RESIDUAL ANALYSIS

Tests for Autocorrelation

LM(1): ChiSqr(16) = 21.79 [0.15]
LM(2): ChiSqr(16) = 8.82 [0.92]

LM(3): ChiSqr(16) = 14.99 [0.53]
 LM(4): ChiSqr(16) = 10.30 [0.85]

Test for Normality: ChiSqr(8) = 34.01 [0.00]

Test for ARCH:

LM(1): ChiSqr(100) = 190.19 [0.00]
 LM(2): ChiSqr(200) = 335.20 [0.00]
 LM(3): ChiSqr(300) = 458.14 [0.00]
 LM(4): ChiSqr(400) = 570.58 [0.00]

Univariate Statistics

	Mean	Std.Dev	Skewness	Kurtosis	Maximum	Minimum
DLYBL	0.00	0.08	-0.15	3.86	0.29	-0.26
DCBR	0.00	1.04	-0.02	3.16	2.86	-3.07
DCDR	0.00	2.45	0.65	4.01	7.79	-6.15
DCMR	0.00	0.59	0.65	3.88	2.15	-1.38

	ARCH(4)	Normality	R-Squared
DLYBL	8.98 [0.06]	7.45 [0.02]	0.29
DCBR	4.76 [0.31]	0.79 [0.67]	0.74
DCDR	13.19 [0.01]	12.72 [0.00]	0.64
DCMR	12.13 [0.02]	12.95 [0.00]	0.51

Appendix C: The short run matrices for the structure with $c_1 = 0$

The coefficient, \tilde{a}_3 was set to zero as it was insignificant (see 4.2). The "Gamma (i)" corresponds to Γ_i in the usual ECM form (equation 28), and t -values are in the brackets. DD16 is the difference of D_t^P , the DD16XXS_{i} denotes the i 'th difference of the shift dummy, Ds_t^{16XX} . In each line DLYBL is the first difference of w_t , CBR, CDR and CMR denote the crude - birth, death and marriage rate respectively ($= b_t \times 1000$, etc.).

LAGGED DIFFERENCES:

GAMMA(1)

	DLYBL{1}	DCBR{1}	DCDR{1}	DCMR{1}
DLYBL	-0.13	-0.00	-0.00	-0.02
	(-1.41)	(-0.01)	(-0.08)	(-2.21)
DCBR	6.37	-0.21	-0.09	0.45

	(6.21)	(-2.96)	(-3.20)	(3.49)
DCDR	-3.48	-0.32	0.03	-1.05
	(-1.44)	(-1.94)	(0.43)	(-3.44)
DCMR	0.41	0.11	0.02	-0.11
	(0.72)	(2.88)	(1.53)	(-1.48)

GAMMA(2)

	DLYBL{2}	DCBR{2}	DCDR{2}	DCMR{2}
DLYBL	-0.14	0.00	0.00	-0.02
	(-1.57)	(0.03)	(1.80)	(-1.45)
DCBR	5.17	-0.08	-0.04	-0.45
	(5.02)	(-1.64)	(-1.64)	(-3.64)
DCDR	-0.62	-0.28	0.05	0.27
	(-0.26)	(-2.42)	(0.86)	(0.93)
DCMR	-1.20	0.02	0.04	-0.10
	(-2.10)	(0.64)	(2.65)	(-1.53)

DUMMIES:

	DD16	DD1643S_{0}	DD1659s_{0}
DLYBL	-0.01	0.02	-0.07
	(-0.26)	(0.20)	(-0.74)
DCBR	1.47	0.88	-3.08
	(2.82)	(0.79)	(-2.69)
DCDR	13.35	5.19	-8.94
	(10.89)	(1.98)	(-3.31)
DCMR	-0.68	-2.89	-0.76
	(-2.34)	(-4.68)	(-1.19)

	DD1643S_{1}	DD1643S_{2}	DD1659S_{1}	DD1659S_{2}
DLYBL	-0.07	-0.14	-0.01	-0.15
	(-0.64)	(-1.39)	(-0.11)	(-1.42)
DCBR	1.14	2.27	0.03	1.17
	(0.95)	(1.93)	(0.02)	(0.94)
DCDR	0.34	-3.29	-6.45	0.62
	(0.12)	(-1.19)	(-2.22)	(0.21)
DCMR	-1.53	1.03	-1.79	-1.16
	(-2.31)	(1.59)	(-2.62)	(-1.68)

References

- Aiyar, S., C.-J. Dalgaard, and O. Moav (2008, June). Technological progress and regress in pre-industrial times. *Journal of Economic Growth* 13(2), 125–144.
- Allen, R. C. (2001). The great divergence in European wages and prices from the middle ages to the first world war. *Explorations in Economic History* 38(4), 411–447.
- Bailey, R. E. and M. J. Chambers (1993). Long-Term Demographic Interactions in Precensus England. *Journal of the royal statistical society, Series A* 156(3), 339–362.
- Boserup, E. (1965). *The Conditions of Agricultural Growth*. London: G. Allen and Unwin.
- Clark, G. (2005). The condition of the working-class in England, 1209-2004. *Journal of Political Economy* 113(6), 1307–1340.
- Clark, G. (2007). *A Farewell to Alms: A Brief Economic History of the World*. Princeton: Princeton University Press.
- Collard, D. (2001). Malthus, Population, and the Generational Bargain. *History of Political Economy* 33(4), 697–716.
- Crafts, N. and T. C. Mills (2007). From malthus to solow: How did the malthusian economy really evolve? *Journal of Macroeconomics (forthcoming)*.
- Dennis, J. G., H. Hansen, and K. Juselius (2006). *CATS in RATS. Cointegration analysis of time series, Version 2*. Evanston, Illinois, USA: Estima.
- Franses, P. H. and N. Haldrup (1994). The effect of additive outliers on tests for unit roots and cointegration. *Journal of Business and Economic Statistics* 12(4), 471–78.
- Galor, O. (2005, February). From Stagnation to Growth: Unified Growth Theory. In P. Aghion and S. Durlauf (Eds.), *Handbook of Economic Growth*, Volume 1, Chapter 4, pp. 171–293. Elsevier.
- Johansen, S. (1992, June). Testing weak exogeneity and the order of cointegration in uk money demand data. *Journal of Policy Modeling* 14(3), 313–334.
- Johansen, S. (1996). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford: Advanced Texts in Econometrics, Oxford University Press.
- Johansen, S. (2006). Confronting the economic model with the data. In D. Colander (Ed.), *Post Warasian Macroeconomics: Beyond the Dynamic Stochastic General Equilibrium Model*, Volume 1, Chapter 15, pp. 287–300. Cambridge University Press.
- Juselius, K. (2006). *The cointegrated VAR model: Econometric methodology and macroeconomics applications*. Oxford University Press.
- Lee, R. (1973, November). Population in Pre-industrial England: An Econometric Analysis. *The Quarterly Journal of Economics* 87(4), 581–607.
- Lee, R. (1977). Methods and Models for Analyzing Historical Series of Births, Deaths and Marriages. In R. Lee (Ed.), *Population Patterns in the Past*, Volume 1, pp. 337–370. Academic Press.
- Lee, R. (1985a). Inverse projection and back projection: A critical appraisal, and comparative results for England, 1539 to 1871. *Population Studies* 39(4), 233–248.
- Lee, R. (1985b, Spring). Population Homeostasis and English Demographic History. *Journal of Interdisciplinary History* 15(4), 635–660.
- Lee, R. (1987, November). Population dynamics of humans and other animals. *Demography* 24(4), 443–466.

- Lee, R. (1993a, January). Accidental and Systematic Change in Population History: Homeostasis in a Stochastic Setting. *Explorations in Economic History* 30(1), 1–30.
- Lee, R. (1993b). A critical assessment of new methods. In D. Reher and R. Schofield (Eds.), *Old and New Methods in Historical Demography*, pp. 7–28. Oxford University Press.
- Lee, R. (1997). Population dynamics: Equilibrium, disequilibrium, and consequences of fluctuations. In M. R. Rosenzweig and O. Stark (Eds.), *Handbook of Population and Family Economics*, Volume 1 of *Handbook of Population and Family Economics*, Chapter 19, pp. 1063–1115. Elsevier.
- Lee, R. and M. Anderson (2002). Malthus in state space: Macro economic-demographic relations in English history, 1540 to 1870. *Journal of Population Economics* 15(2), 195–220.
- Lin Lee, M. and D. Loschky (1987, September). Malthusian Population Oscillations. *The Economic Journal* 97(387), 727–739.
- Malthus, T. R. (1798). *An Essay on the Principle of Population*. London.
- Miller, M. H. and C. W. Upton (1986). *Macroeconomics: A Neoclassical Introduction*. The University of Chicago Press.
- Muth, J. F. (1961). Rational Expectations and the Theory of Price Movements. *Econometrica* 29(3), 315–335.
- Møller, N. F. (2008). Bridging economic theory models and the cointegrated vector autoregressive model. *Economics: The Open-Access, Open-Assessment E-Journal* 2(2008-21).
- Nicolini, E. A. (2007). Was Malthus right? A VAR analysis of Economic and Demographic Interactions in Pre-industrial England. *European Review of Economic History*, II, 99–121.
- Nielsen, H. B. (2008). Influential Observations in Cointegrated VAR models: Danish Money Demand 1973-2003. *The Econometrics Journal* 11(1), 1–19.
- Persson, K. (1988). *Pre-industrial economic growth: Social organization and technological progress in Europe*. Oxford: Blackwell.
- Rahbek, A., E. Hansen, and J. G. Dennis (2002). ARCH innovations and their impact on cointegration rank testing. Preprint no.12, 1998, Department of Theoretical Statistics. Working paper no.22, Centre for Analytical Finance.
- Schultz, T. P. (1981). *Economics of population*. Addison-Wesley Publishing Company.
- Sharp, P. and J. Weisdorf (2008). From Preventive to Permissive Checks: The changing nature of the Malthusian relationship between nuptiality and the price of provisions in the nineteenth century. Forthcoming in *Cliometrica*.
- Simon, J. L. (1977). *The Economics of Population Growth*. Princeton: Princeton University Press.
- Wrigley, E. A. and Schofield (1989). *The Population History of England 1541-1871: A Reconstruction*. Cambridge University Press.