HEDG
HEALTH, ECONOMETRICS AND DATA GROUP
THE UNIVERSITY of York

WP 14/07

A synthesis of the Grossman and Becker-Murphy models of health and addiction: theoretical and empirical implications

Andrew M. Jones; Audrey Laporte; Nigel Rice &
Eugenio Zucchelli

April 2014

york.ac.uk/res/herc/hedgwp

A synthesis of the Grossman and Becker-Murphy models of health and addiction: theoretical and empirical implications

Andrew M. Jones [a] Audrey Laporte [b,c] Nigel Rice [a,d,*]
Eugenio Zucchelli [d]

[a] *Department of Economics and Related Studies, University of York, YO10 5DD, UK*
[b] *Institute for Health Policy, Management and Evaluation, University of Toronto, ON M5T 3MS, Canada*
[c] *Canadian Centre for Health Economics, University of Toronto, Canada*
[d] *Centre for Health Economics, University of York, UK*

[d] *Division of Health Research, Lancaster University, LA1 4YG, UK*

April 2, 2014

Summary

This paper develops and estimates a model that integrates two fundamental theories of individual health behaviour: the Becker-Murphy model of rational addiction and the Grossman model of health investment. We define an individual's lifetime smoking consumption pattern and investments in health capital as simultaneous choices within a single optimisation problem allowing for the presence of an addiction stock and investments in preventive medical care. The resulting system of first-order difference equations is reduced to a single fourth-order difference equation defined both for smoking and health and which preserve the dynamic roots of the system. GMM systems estimation using the British Household Panel Survey reveals strong persistence in the evolution of both smoking consumption and health capital with direct effects of past health and smoking observed for up to three and four lagged periods for men and women respectively. Conditional on dynamics there is a limited role for the direct effects of socio-economic status. A convincing understanding of an individual's optimal lifetime health trajectory requires an appreciation of how both investments and accumulated disinvestments in health separately impact on the dynamics of health capital. The integrated approach presented here offers such a framework.

JEL classification: C1, C3, C5, C6, I1.

Key words: Addiction; Health capital; Smoking; Dynamic panel data models

---

*Corresponding author: Nigel Rice. Tel.: +44 1904 321411.
E-mail address: nigel.rice@york.ac.uk

1 Introduction

The dominant theoretical model in the economics of health broadly speaking is the Grossman model (Grossman, 1972) which extends the theory of human capital to the demand for health and medical care over the life-cycle and provides the foundation for a large body of empirical research. The dominant theoretical model in the economic analysis of the consumption of addictive commodities is the Becker-Murphy model of rational addiction (Becker and Murphy, 1988; henceforth B-M). Each is a subset of the general model of individual-level investment in human capital concerned with an individual's inter-temporal investment problem. To date, the two models have tended to be treated as separate entities, both for theoretical and empirical analyses. In this paper we treat the two theories as complementary and integrate them into a single framework. Grossman's model of investment in health capital considers the way an individual's behaviour at one point in time affects their health over the entirety of their planning horizon. The model is usually set up such that investment in health is a 'good' in the sense that it has a positive marginal product in the production function for health. The consumption of addictive 'bads' such as smoking, however, can be modelled using the same conceptual framework (but with a negative marginal product in the production function) and 'good' and 'bad' investments considered simultaneously.

We develop and estimate a model that explicitly integrates the dynamics of addiction and the human capital model of health investment. We adopt an optimal control approach and model smoking behaviour and positive investment in health capital as simultaneous choices of a single optimization problem, allowing for the presence of an addiction stock together with a stock of health capital and investments in preventive medical care. We are not the first authors to include both 'goods' and 'bads' in a Grossman framework (see Forster, 2001, for example). The novelty in this paper is to construct the integrated model in a manner that is consistent with the medical literature on the health effects of smoking (e.g. Doll et al., 2004). That literature suggests that smoking has a cumulative effect in the sense that its impact on an individual's health capital depends not on how much they are smoking today but on the amount of smoking damage they have accumulated through their lifetime. Consistent with the B-M framework we model smoking damage as a state variable in an optimal control framework allowing it to have its own intrinsic dynamics and consistent with the Grossman framework we also model health capital as a state variable with its own intrinsic dynamics, but enter the stock of smoking damage as an explanatory variable in the equation explaining health damage. This allows us to let smoking not only have its own dynamics but to have a different time profile of effect on health than other health related activities.

The resulting inter-temporal optimisation leads to a system of interrelated first-order difference equations. Empirical implementation of this system is

complicated by the fact, not uncommon in health datasets, that we lack data on certain key choice variables - most obviously on cumulative smoking damage. We therefore reduce the system from four inter-related first-order difference equations to a pair of stand alone fourth-order difference equations, one in health and one in cigarette consumption. An implication of the derivation is that the roots of each of the single fourth-order difference equations are the same as the four roots of the original system of four first-order difference equations and accordingly can be used to investigate the dynamics of the system.

This is the first attempt, as far as we are aware, to formally reconcile the two fundamental economic theories of individual health behaviour, the human capital model of health investment and the model of rational addiction (RA), into a single framework. Secondly, the integrated framework explicitly allows health capital and smoking damage to have distinct dynamic shadow prices which is not the case when the current level of cigarette consumption is simply entered directly as an input into the health capital production function. The differing evolution of these two dynamic shadow prices can be seen as underlying the way different individuals choose different lifetime pathways for health and addiction capital. Thirdly, we present and discuss an empirical implementation of the combined model in a continuous choice framework. This requires the estimation of dynamic panel data models involving fourth-order lags of the dependent variable. We do this using a generalised method of moments systems estimator deriving instruments for the lags of the dependent variable from past period observations. This necessitates the use of rich and mature panel data and we make use of 18 waves of the British Household Panel Survey (BHPS).

Our estimates confirm the strong persistence of both smoking consumption and health capital with direct effects on current health and smoking observed for up to three and four lagged periods for men and women respectively. The dominant real roots which drive the long-term behaviour of health and smoking consumption are both postive and less than one in absolute value, and are numerically similar across the equations and generally also for men and women. Conditional on the intrinsic dynamics in the health and smoking equations, we find that for men higher household income is associated with being healthier while for women health appears independent of income. The effect for men, however, is small and appears to corroborate findings elsewhere in the literature on the relationship between income and health (e.g. Contoyannis et al., 2004; Frijters et al., 2005; Meer, 2003). Education while generally displaying the expected gradients for health (positive) for women, is not statistically significant. Education does, however, display a more pronounced gradient for smoking with effects greater (and statistically significant) for women compared to men. Non-white ethnicity is significantly associated with lower daily smoking for women. There is evidence of geographical variation, particularly for women where areas generally associated with increased levels of deprivation

(compared to the baseline of South East of England) are associated with decreased health and increased smoking.

The paper is organised as follows. Section 2 provides a background to the B-M model of RA and the Grossman model of health capital investment and maps out the way in which these theories can be integrated. Section 3 presents our theoretical approach that nests the Grossman model within the RA framework. Sections 4 and 5 present our empirical approach and data respectively. Results follow in section 6 and section 7 presents a discussion of the findings.

2 Background

Early applications of the RA model were focused on the analysis of cigarette consumption and were based on both aggregated (Becker, Grossman, Murphy, 1994; Keeler et al., 1993) and individual-level data (Chaloupka, 1991). In this framework, rationality involves forward-looking behaviour or a plan to maximise utility over time and accordingly individuals anticipate the future (harmful) consequences of their current choices. These studies appear to broadly support the main implications of rational addictive behaviour, and reject myopic behaviour.[1] In standard RA models addiction is often considered the only factor that affects an individual's health while health capital and its evolution over the life-cycle are not explicitly modelled.

More recent applications on tobacco consumption employ dynamic panel data specifications and focus on issues such as errors-in-variables, data censoring and individual-level unobserved heterogeneity (e.g. Jones and Labeaga, 2003; Labeaga, 1999). In general, these models also reject myopic behaviour and support RA. The dynamic framework of addiction has also been applied to the analysis of consumption of other addictive goods such as alcohol (Baltagi and Griffin, 2002; Grossman, 1993; Grossman, Chaloupka, and Sirtalan, 1998; Waters and Sloan, 1995) and illicit drugs such as cocaine, heroin and marijuana (e.g. Grossman and Chalopka, 1998; Saffer and Chaloupka, 1999).

The seminal work of Grossman (Grossman, 1972) draws from the literature on human capital theory (Becker, 1965; Becker and Woytinsky, 1967; Mincer, 1974) and applies it to the demand for health and medical care over the life-cycle. In this framework, individuals inherit an initial amount of health capital stock that, while depreciating with age, can be increased through investments in health (for example, via medical care). Extensions to the model tend to focus on its underlying assumptions and implications concerning optimal investments (e.g. Ehrlich and Chuma, 1990).

[1]Note, however, that most studies are not conclusive in this respect and often produce implausible estimates of discount rates. However, see Gruber and Koszegi (2001) for a discussion of potential dynamic inconsistencies in preferences with respect to smoking. Note that our paper is concerned with embedding the Grossman model within the RA framework and is not explicitly concerned with estimating discount rates.

There are few economic studies that model health and smoking jointly. Adda and Lechene (2013) employ hazard models on data drawn from the Swedish Survey of Living Conditions merged with death records to analyse the effect of smoking on mortality. They find evidence of selection into smoking such that individuals with poorer health are more likely to smoke and that the effect of smoking on mortality appears to be larger for individuals with a potentially longer life expectancy as measured by a series of proxies for health status. Balia and Jones (2008) estimate a recursive system of equations for lifestyles, morbidity and mortality and explore health inequalities in mortality using decomposition techniques on data from the British Health and Lifestyle Survey. They find that lifestyles appear to contribute strongly to inequality in mortality, reducing the direct role of socio-economic status. Darden (2012) proposes a RA model of smoking augmented by a Bayesian learning process through which individuals acquire information about their own health and use it to make decisions about smoking. Therefore, these studies have focused on specific aspects of the interactions between health (longevity and individual health information) and lifestyles but do not attempt to formally set out a model of the dynamic interrelations between smoking and health capital.

Integrating investment in health capital into the B-M framework poses fundamental challenges. First is the difficulty in defining and solving a two state variable dynamic optimisation problem - with one state variable representing the accumulated stock of health and the other the accumulated stock of addiction (e.g. Leonard and Long, 1992; Pitchford, 1977). Secondly, estimation of the dynamic models which this general theoretical perspective implies, requires long panels of individual level data. It is standard in the applied literature on smoking behaviour to include a variable representing current state of health, and, conditional on data availability, variables on past health shocks (e.g. Arcidiacono, 2007). Nonetheless very few papers treat the level of health as a variable affected by smoking behaviour while allowing smoking to be dynamically endogenous, and the time paths of smoking and health as part of the same inter-temporal optimization problem. In this paper we adopt a multi-state variable optimal control approach to combine the Grossman and B-M models as it appears to be the most promising approach for understanding the nature of the lifetime trajectory of individuals' health capital stocks. Our view is that in combining these two models, we need to recognize that smoking behaviour and positive investment in health capital are simultaneous choices emerging out of a single optimization problem.

3 Theoretical model

3.1 The standard model of rational addiction

The RA model of B-M analyses the smoking choices in an inter-temporal optimization framework where the quantity of cigarettes smoked is the control

variable and the stock of addiction capital is the state variable. In this model, addiction is generally reflected in the assumption that, as the stock of addiction increases, a consumer's preferences are changed increasingly to favour the consumption of cigarettes (via reinforcement). This is despite the assumption that the individual also derives disutility from their stock of addiction capital due to the detrimental effects of smoking on health. In discrete time terms, the elements of the standard RA model can be written as:[2]

$$\max \sum_{t=0}^{\infty} \beta^t U\left(X_t, S_t, A_t\right), \tag{1}$$

$$\text{with } U_X > 0, U_{XX} < 0; U_S > 0, U_{SS} < 0; U_A < 0, U_{AA} < 0; U_{SA} > 0;$$

$$A_t = f\left(S_t\right) + \left(1 - \delta_A\right) A_{t-1}, \quad f_S > 0, f_{SS} < 0; \tag{2}$$

$$Y_t = X_t + p_S S_t, \tag{3}$$

where S_t is the quantity of cigarettes consumed in period t, A_t is the stock of addiction, X_t refers to other commodities and Y_t is income. The relative price of S is p_S with the price of X normalized to 1; δ_A is the rate at which the stock of addiction decays and β is the discount factor. The marginal utilities of X_t and S_t are positive and decreasing while that of A_t is negative and decreasing. The intertemporal utility function (1) is maximized by choice of X_t and S_t. [3] The equation of motion for the addiction stock is given by (2). Here smoking is assumed to have a positive effect on A_t, although due to the difficulty in measuring A_t, it is usually assumed that $f(S_t)$ is linear for A_t.[4] This problem is solved as an optimal control problem, yielding necessary conditions which include the equation of motion of A_t and a terminal transversality condition and an Euler equation for S_t. The resulting first-order difference equation in S_t interacts with the first-order difference equation for A_t. This is a straightforward optimal control problem, which could be analysed qualitatively using a phase diagram in S_t and A_t.

A standard problem in the empirical literature is the difficulty of measuring A_t.[5] In the majority of applications this is handled by the reduction of the system of two first-order difference equations, one in S_t and one in A_t, to a single second-order difference equation in S_t, where S_t (consumption of cigarettes) is the observable variable. This reduction is standard in the dynamics literature (see Ferguson and Lim, 2003), and preserves the characteristic

[2]RA models often include a wealth equation (e.g. Becker and Murphy, 1988). This is omitted here as it is not essential to our narrative.

[3]In this version of the model, where borrowing is not permitted, X_t and S_t are tied together by the budget constraint (3) which allows us to substitute X_t out of the problem.

[4]This assumption of linearity does not impact on the qualitative solution to the problem.

[5]Some studies have used biological markers which are assumed to reflect the addictive stock (Adda and Cornaglia, 2006, 2010).

roots of the system, which drive the dynamics of both A_t and S_t. In what follows we will employ a similar strategy and reduce a system of four first-order difference equations resulting from optimization into two single fourth-order difference equations; one in health and one in smoking, while preserving the characteristic roots of the initial system.

3.2 Integrating the Grossman model

A limitation of the standard B-M model is that the only factor which represents or affects an individual's health is the stock of addiction capital, A_t. We augment the model with a Grossman model of investment in health capital, where the investment goods include harmful health 'bads' (cigarette consumption) as well as the usual investment 'goods'. In other words, the rational, lifetime investor in health capital should consider the range of factors that affect his lifetime health simultaneously, and this should be reflected in theoretical models as well as their empirical counterparts.

We extend the B-M model to include the demand for health. We write the individual's lifetime utility function as:

$$\sum_{t=0}^{\infty} \beta^t U\left(X_t, S_t, H_t, A_t\right),$$

with $U_X > 0, U_{XX} < 0; U_S > 0, U_{SS} < 0; U_A < 0, U_{AA} < 0; U_{SA} > 0; U_H > 0, U_{HH} < 0$. The variable, H_t, is the stock of health capital at time t. Addiction capital, A_t, remains as an argument in the utility function since an individual derives disutility from being addicted, even though the health effects of smoking, which rises with the cumulative amount smoked over time, could be absorbed into H_t. H_t cannot be purchased directly, but rather must be produced using an input M_t, which can be purchased. Accordingly, the budget constraint is now,

$$Y_t = X_t + p_S S_t + p_M M_t,$$

where the price of cigarettes, p_S, and medical care, p_M, are relative to the price of other consumption goods, X.

We retain the equation of motion for A, given by (2), and specify the following equation of motion for H,

$$H_{t+1} = \left[1 - \delta_H\right] H_t + h\left(M_t\right) + l\left(A_t\right), \qquad h_M > 0, h_{MM} < 0; l_A < 0, l_{AA} < 0.$$

A appears in the equation of motion for H to represent the cumulative effects of smoking on cumulated health capital. Even if an individual quits smoking such that $S = 0$, the accumulated addiction stock will continue to harm their health (reduce health capital) until it has decayed away. We assume that investing in health via the purchase of medical care, M_t, does not produce

utility per se.[6] Consumption which is neutral or bad for your health does produce utility: $l(0) > 0$, $l_A(A) < 0$. Hence, $l(0)$ represents an extra health increment obtained by someone who has accumulated no addiction capital, and allows us to define a small improvement in the health of someone who quits smoking. This accords with results from the biomedical literature suggesting that if an individual quits smoking relatively early in life, before their stock of addiction has created permanent health damage, then their survival probability should revert to that of a never smoker in a relatively short time period (Doll et al., 2004). Conversely, someone who has accumulated a large addiction stock might not be able to benefit from quitting in the same way as individuals with a smaller accumulated stock, so that for most long term smokers, $l(A)$ will be negative. After substituting X out of the utility function, using the budget constraint, the problem, written in Chow's Lagrange multiplier format (see Ferguson and Lim, 2003) is:

$$Max \sum_{t=0}^{\infty} \beta_t U\left(Y_t - p_S S_t - p_M M_t, S_t, A_t, H_t\right) - \beta_{t+1}\lambda_{t+1}\left[H_{t+1} - \left[1-\delta_H\right]H_t \right.$$
$$\left. -h\left(M_t\right) - l\left(A_t\right)\right] - \beta_{t+1}\mu_{t+1}\left[A_{t+1} - \left[1-\delta_A\right]A_t - f\left(S_t\right)\right],$$

where λ and μ represent the shadow price of health and addiction capital respectively. Since addiction is a 'bad', μ is negative while λ is positive because health capital is a 'good'.

The corresponding first-order conditions are:

$$\beta\mu_{t+1} = p_S U_X(t) - U_S(t), \tag{4}$$

$$\beta\lambda_{t+1} h\left(M_t\right) = p_M U_X(t),$$

$$\beta\left[1-\delta_H\right]\lambda_{t+1} = \lambda_t - U_H(t),$$

$$\beta\left[1-\delta_A\right]\mu_{t+1} = \mu_t - U_A(t) - \beta\lambda_{t+1} l\left(A_t\right). \tag{5}$$

The equation of motion for the shadow price of addiction capital (5) now contains the shadow price of health capital, λ, reflecting the fact that the stock of addiction capital is a determinant of the stock of health capital. The first-order conditions can be rearranged to eliminate the Lagrange multipliers. In so doing, we assume that these conditions always hold, meaning that individuals

[6]More generally M can represent any good which is beneficial for health but yields no direct utility. We follow the standard approach in the literature and denote this as medical care.

are in fact optimizing. Rearranging yields the following system of four first-order difference equations in M_t, S_t, A_t and H_t:

$$[1-\delta_H]\frac{(p_M U_X(t))}{h(M_t)} = \frac{p_M U_X(t-1)}{\beta h(M_{t-1})} - U_H(H_t), \tag{6}$$

$$[1-\delta_A]\,p_S U_X(t) - [1-\delta_A]\,U_S(t) = p_S U_X(t-1) - \frac{U_S(t-1)}{\beta} - U_A(t) - \frac{p_M U_c(t)}{h(M_t)\,l(A_t)},$$

$$A_{t+1} = [1-\delta_A]\,A_t + f(S_t),$$

$$H_{t+1} = [1-\delta_H]\,H_t + h(M_t) + l(A_t).$$

This system of first-order difference equations can be translated into four linearized first-order difference equations in H_t, S_t, A_t and M_t which, in turn, can be reduced into a single fourth-order equation in S_t or H_t. These form the basis of our empirical model.

4 Empirical models

4.1 Derivation of estimating equations

The starting point for our estimation strategy follows the original B-M model as described by equations (1) to (3). The first-order conditions for this model can be rearranged as a pair of first-order difference equations in S_t and A_t. However, the B-M model is typically estimated as the following linear second-order difference equation in S:

$$S_t = \alpha_0 + \alpha_1 S_{t+1} + \alpha_2 S_{t-1} + \epsilon_t. \tag{7}$$

Typically this is referred to as a forward looking second-order difference equation, since it contains S_{t+1} on the right-hand side and it is often said that the forward looking nature of this equation reflects the rationality of the consumption decision. In fact, rationality is a consequence of this equation having emerged from an inter-temporal optimization problem and is in many ways inherent in the first-order condition (4) which we can rewrite as: $-\beta\mu_{t+1} = U_S(t) - p_S U_X(t)$. In this expression, μ_{t+1} is the shadow price of another unit of addiction capital (negative since addiction capital yields disutility). The right hand side of this expression is the net benefit in utility terms that the consumer derives from consuming another unit of S_t net of the utility given up because consuming more S requires consuming less X. The fact that the benefit is derived in period t and the cost in $t+1$ is the essence of the forward-looking nature of the decision.

The standard empirical specification of the RA model (7) is written with one lead and one lag of the dependent variable on the right hand side. This specification, is however, not necessary and could be rearranged and written with two lags on the right hand side. The RA problem is, as noted above, an inter-temporal optimization problem which is typically set up as an optimal control problem. The solution equations to an optimal control problem are necessary conditions for optimizing the present value of the stream of future utilities which will arise from future consumption decisions, taking account of how current consumption affects future addiction. The process of reducing the two interrelated first-order difference equations which fall out of the necessary conditions of the usual version of the RA model to a single second-order difference equation does not affect the fact that the necessary conditions are forward looking, regardless of how we happen to write the second-order equation.

The fact that (7) is a single second-order difference equation in S_t rather than a pair of first-order difference equations, one in S_t and one in A_t, is a reflection of the result that a pair of interrelated first-order difference equations can be reduced to a single second-order difference equation in either of the variables by suitable substitutions. In the case of the standard B-M reduction underlying the typical estimation equation is the assumption that the system of first-order conditions has first been linearised and then the necessary substitutions made. One implication of the reduction process is that the roots of the single second-order equation are the same as the roots of the system of two first-order difference equations. This means that the dynamics of S_t derived from the estimated single second-order system will be the same as the dynamics that would be observed for S_t were it possible to estimate the original pair of difference equations. It is not possible to estimate the original pair of equations since A_t is unobservable, so this result means that in estimating a single second-order difference equation for S_t, while we lose some information about A_t, but the dynamics of S_t in the sense of its behaviour over time are the same as would be found were it possible to estimate the original system.

Thus the estimation of a second-order difference equation for S_t is not so much a sign of rationality as it is a matter of making sure that we extract as much information about the intrinsic time path of S_t which, since the consumer is assumed to be rational, depends on the time path of A_t. Rationality is inherent in the optimal control model we are using to explain an individual's lifetime path of cigarette consumption.

Our estimation strategy set out below is based on an extension of the B-M approach, where we reduce our set of four interrelated first-order difference equations (6) to a single fourth-order difference equation. For tractability, we follow a practice which is common in the literature on analytical dynamics, when expressions for characteristic roots are being found for a model which is intrinsically nonlinear, and assume that we are working with a linear approximation to the original non-linear system. These assumptions about functional form lead to the following four linearised first-order difference equations:

$$
\begin{aligned}
H_t &= \vartheta_0 + \vartheta_1 H_{t-1} + \vartheta_2 S_{t-1} + \vartheta_3 A_{t-1} + \vartheta_4 M_{t-1} + \vartheta_5 X_t + \eta_t^h \\
S_t &= \alpha_0 + \alpha_1 H_{t-1} + \alpha_2 S_{t-1} + \alpha_3 A_{t-1} + \alpha_4 M_{t-1} + \alpha_5 X_t + \eta_t^s \\
A_t &= \gamma_0 + \gamma_1 H_{t-1} + \gamma_2 S_{t-1} + \gamma_3 A_{t-1} + \gamma_4 M_{t-1} + \gamma_5 X_t + \eta_t^a \\
M_t &= \omega_0 + \omega_1 H_{t-1} + \omega_2 S_{t-1} + \omega_3 A_{t-1} + \omega_4 M_{t-1} + \omega_5 X_t + \eta_t^m \qquad (8)
\end{aligned}
$$

These four first-order difference equations can be reduced into single fourth-order linear equations for estimation (see Appendix A for details). We do this separately for smoking and health which yields the following general forms:

$$
\begin{aligned}
&S_{it} = \phi_{s0} + \phi_{s1} S_{it-1} + \phi_{s2} S_{it-2} + \phi_{s3} S_{it-3} + \phi_{s4} S_{it-4} + \theta_{s1} X_{it}^s + \theta_{s2} X_{it-1}^s \\
&+\theta_{s3} X_{it-2}^s + \theta_{s4} X_{it-3}^s + \psi_s W_i^s + \mu_{is} + (\epsilon_{it} + \rho_{\epsilon 1}\epsilon_{it-1} + \rho_{\epsilon 2}\epsilon_{it-2} + \rho_{\epsilon 3}\epsilon_{it-3}) \qquad (9)
\end{aligned}
$$

$$
\begin{aligned}
&H_{it} = \phi_{h0} + \phi_{h1} H_{it-1} + \phi_{h2} H_{it-2} + \phi_{h3} H_{it-3} + \phi_{h4} H_{it-4} + \theta_{h1} X_{it}^h + \theta_{h2} X_{it-1}^h \\
&+\theta_{h3} X_{it-2}^h + \theta_{h4} X_{it-3}^h + \psi_h W_i^h + \mu_{ih} + (\varepsilon_{it} + \rho_{\epsilon 1}\varepsilon_{it-1} + \rho_{\epsilon 2}\varepsilon_{it-2} + \rho_{\epsilon 3}\varepsilon_{it-3}) \qquad (10)
\end{aligned}
$$

for $i = 1, \ldots, N$, and $t = 1, \ldots, T$.

Here S_{it} represents smoking consumption for individual i at time t; H_{it} is the stock of health; X_{it}^s and X_{it}^h are sets of exogenous time-varying predictors of smoking consumption and health respectively. W_i^s and W_i^h, and μ_{is} and μ_{ih} are respectively time-invariant predictors and time-invariant individual-specific unobserved effects for smoking and health. We assume that $\epsilon_{it} \sim i.i.d.\,(0, \sigma_\epsilon^2)$ with $E(\epsilon_{it}) = 0$ and, similarly, $\varepsilon_{it} \sim i.i.d.\,(0, \sigma_\varepsilon^2)$ with $E(\varepsilon_{it}) = 0$. Also, we assume that $E(X_{ir}^s, \epsilon_{it}) = 0, E\left(X_{ir}^h, \varepsilon_{it}\right) = 0$ for $\forall r, t$. That is, X^s and X^h include strictly exogenous regressors uncorrelated with ϵ and ε respectively. Both sets of time-varying predictors may, however, be correlated with their respective unobserved effects, μ_{is} and μ_{ih}. We assume that the time-invariant regressors W_i^s and W_i^h are orthogonal to the unobserved time-varying effects.[7]. Note that the equations contain third-order moving average processes in the error (MA(3)).

[7]It might be argued that this assumption is untenable. However, relaxing the assumption requires either the identification of external instruments for W_i, or relying on instruments internal to the model. The latter might consist of transformations of $X_{it}, \ldots, X_{it-3}$ where a subset of these are assumed to be uncorrelated with the unobserved individual specific effect in the spirit of Hausman and Taylor type estimators (Hausman and Taylor, 1981). Alternatively, differences in the lags of the dependent variable can be used as instruments. It is not surprising, however, that such instruments are weak when used in this context. Estimates of the lagged dependent variables, $S_{it-1}, \ldots, S_{it-4}$ and $H_{it-1}, \ldots, H_{it-4}$ do not change dramatically for models estimated without the vector of time-invariant regressors.

4.2 Estimation

It is well known that OLS estimation of dynamic panel data models with fixed T is biased such that the parameters ϕ_{s1} , ϕ_{s2}, ϕ_{s3} and ϕ_{s4} in (9) and ϕ_{h1} , ϕ_{h2}, ϕ_{h3} and ϕ_{h4} in (10) will be overestimated (Nickell, 1981). Standard fixed effects estimation is downwardly biased. Instead estimation by the generalized method of moments (GMM) is favoured, constructing instruments for the lagged dependent variables, $(S_{it-1}, S_{it-2}, S_{it-3}, S_{it-4})$ and $(H_{it-1}, H_{it-2}, H_{it-3}, H_{it-4})$ from past values of the regressors (Arellano and Bond, 1991). The standard approach to dealing with the individual fixed effect (μ_{ih} and μ_{is}) is to first-difference the model. The additional serial correlation induced in the first-differenced error term informs the moment restriction imposed on the model and the choice of instruments. In the standard autoregressive dynamic panel data model with a single lag of the dependent variable, no exogenous regressors and no serial correlation in the error term in levels form, values of the dependent variable lagged two periods or more are valid instruments in the equation in first differences. For $T \geq 3$, there are $m = (T-2)(T-1)/2$ moment restrictions of the form $E[Z_i'\bar{\eta}_i] = 0$ where $\bar{\eta}_i = (\bar{\eta}_{i3} \cdots \bar{\eta}_{iT})'$ and $\bar{\eta}_{it} = \eta_{it} - \eta_{it-1}$. These ideas naturally extend to the case of additional lags of the dependent variable and where serial correlation in the error exists in levels.

After first-differencing (9), the moment restrictions can be written in vector form as $E\left(Z_i^{'D}\Delta\epsilon_i\right) = 0$, where $\Delta\epsilon_i = (\Delta\epsilon_{3i}, \ldots, \Delta\epsilon_{iT})'$. $\Delta\epsilon_i = \epsilon_{it} - \epsilon_{it-1}$ and Z_i^D is a block diagonal matrix whose jth block is given as (see Appendix B for details):

$$\begin{aligned} Z_i^D \;=\; & \Big[\operatorname{diag}(S_{i1}, \ldots, S_{is}) \vdots (\Delta X_{i6}, \ldots, \Delta X_{iT})' (\Delta X_{i5}, \ldots, \Delta X_{iT-1})' \\ & (\Delta X_{i4}, \ldots, \Delta X_{iT-2})' (\Delta X_{i3}, \ldots, \Delta X_{iT-3})' \Big] \\ & \text{for } s = 1, \ldots, T-5; T \geq 6 \end{aligned} \tag{11}$$

Accordingly, lags of the levels of the dependent variable form instruments for the difference model whilst the exogenous regressors act as instruments for themselves.

Due to weak instruments, the GMM estimation in first difference form can perform poorly where there exists higher-order autoregressive terms (persistence). Blundell and Bond (1998) suggest the use of a systems estimator that exploits additional moment conditions based on lagged differences of the dependent variable as instruments for a model in levels (also see Blundell, Bond and Windmeijer, 2001). These additional moment conditions are valid under mean stationarity of the initial condition: $E(\mu_{is}\Delta y_{i2}) = 0, \forall i$. Estimation in levels form also allows the identification of the coefficients on the time-invariant regressors, W_i^s and W_i^h. As these are assumed to the uncorrelated with the

individual unobserved effects they act as their own instruments for a model in levels. We follow this approach and estimate the fourth-order difference equations using system GMM. This approach effectively augments the above instrument set with a set of moment conditions $E\left[Z_i'^L \epsilon_i\right] = 0$ where

$$Z_i^L = \left[\text{diag}\left(\Delta S_{it-4}\right) \vdots \left(W_i^s\right)'\right] \quad T \geq 6 \tag{12}$$

Blundell and Bond (1998) show that these additional moment conditions are informative where data are persistent and instruments for the differenced equation are potentially weak, resulting in smaller finite sample bias and increased efficiency.[8]

We begin by estimating the fourth-order difference equations set out in (9) and (10) using a systems GMM estimator with instruments defined in (11) and (12). Due to the length of panel observations available in the BHPS the set of instruments is large which can result in poor performance. Accordingly, we place restrictions on the instrument set to reduce its dimensionality by removing instruments further away from the observation period, t. Efficient two-step estimation applying Windmeijer's finite sample correction to the estimated variance is used (Windmeijer, 2005). Specification tests of autocorrelation and the Sargan test of over-identifying restrictions are computed (see Arellano and Bond, 1991). In addition we compare GMM systems estimation to the within estimator (biased downwards) and OLS (biased upwards).

5 The British Household Panel Survey (BHPS)

5.1 Data and sample

We estimate models on data drawn from 18 waves (1991 - 2009) of the British Household Panel Survey (BHPS). The BHPS is one of the longest and most comprehensive panel surveys currently available. The survey includes individual-level information on demographic and household characteristics; lifestyles including smoking habits; physical and mental health, well-being and the use of health care; education; job histories and interactions with the labour market as well as income and wealth. Its design and main content closely resemble those of other major panel data surveys such as the U.S. Panel Study of Income Dynamics (PSID) and the German Socio-Economic Panel (GSOEP). The BHPS initial sample (wave 1; 1991) consists of 5,500 private households and 10,264 individuals from England, Scotland and Wales.[9] Original sample members are followed as they transit to different households and

[8]Note that further lagged differences of the dependent variable are redundant when combined with instruments for the first-differenced equation (see Blundell and Bond, 1998).

[9]Before 1999 (wave 9), Scottish individuals were only sampled if they resided South of the Caledonian Canal.

interviews are conducted annually to all adult individuals (aged 16 years or over), including new members of the households. Extension samples of around 1,500 households in each of Scotland and Wales and of 2,000 household from Northern Ireland were added in 1999 and 2001 respectively, making the BHPS representative of the whole UK.[10]

Our samples of interest consist of unbalanced panels of individuals who reported the consumption of cigarettes in at least one of the 18 waves of the survey. Never-smokers are excluded as such individuals tell us little about addictive behaviours. Within this sample, individuals for whom we observe sufficient data to estimate the model are included. Given the lag structure of the empirical model and the requirement to construct instruments from prior waves of data, this requires individuals to be observed for at least six consecutive waves.[11] Clearly, responses on individuals for whom we observe non-missing values on the set of variables of interest are included in the model. Accordingly, models for health are estimated on an unbalanced sample of 14,635 observations on 2,315 individuals for men and 17,674 observations on 2,701 individuals for women. Similarly, for smoking the respective samples are 18,407 observations on 2,864 individuals for men and 21,915 observations on 3,340 individuals for women.

5.2 Smoking

The BHPS contains two main self-reported indicators on smoking for adult individuals: smoking status and the daily number of cigarettes smoked. Information on smoking status is based on the question: "Do you smoke cigarettes?" from which we create a dummy variable taking value 1 if the individual is a smoker and 0 otherwise. Since the focus of our analysis is on addiction, our empirical models employ data on individuals who self-report being a smoker at least once during the period of the survey (potential smokers). Accordingly, we exclude individuals who reported not smoking throughout the 18 waves of the BHPS (see Table 1 for basic descriptive statistics for samples of interest). Information on the number of cigarettes smoked is derived from the following question "Approximately how many cigarettes a day do you usually smoke?". While this question is only asked to smokers, 0 is a possible answer that identifies occasional or social smokers (that is, individuals who defined themselves as smokers but report an average daily consumption of 0 cigarettes). Due to the heaping of responses that is typical of self-reported information on the quantity of cigarettes smoked (i.e. large number of responses concentrated at particular levels of smoking consumption), we recode consumption by considering multiples of five cigarettes (we refer to these as 'half packs').

[10] For further details on the BHPS sample structure, see Lynn (2006).

[11] The lag structure imposed by serial correlation in the error-term determines the exact number of lags required to construct valid instruments.

5.3 Health

The BHPS contains a wide range of self-reported categorical variables of individual health status including the General Health Questionnaire (GHQ) on subjective well-being (Goldberg and Williams, 1988) and, in waves 9 and 14, the Short Form 36 (SF-36) health survey. The SF-36 is a standardised health questionnaire including 35 psychometric-validated questions about 8 different dimensions of both physical and mental health (physical functioning; role physical limitations; bodily pain; general health; mental health: vitality; social functioning; role emotional limitations; mental health) (Ware et al., 1993). Each dimension contains a set of items which present respondents with a series of choices about the perception of their own health. Information from all these health questions is used to build a summary measure of health, the SF-36 general index.[12]

We follow Brazier, Roberts and Deverill (2002) and use selected questions from 6 of the original 8 dimensions of the SF-36 (physical functioning; role limitations; social functioning; bodily pain; mental health; vitality) to build a preference-based index measure of health called the SF-6D that is defined on a continuous scale ranging between 0 (an health state equivalent to death) to 1 (full health).[13] More specifically, Brazier et al. employed health information from selected items of these 6 dimensions and combine it with health state utility values to define a utility-based measure of health. Health utility values were retrieved through a preference-based valuation survey of the UK general population. We apply these weights on the items from the 6 dimensions of the SF-36 to generate SF-6D values for individuals in waves 9 and 14 of the BHPS. In order to recover SF-6D values for all individuals in each wave, we regress SF-6D values onto the BHPS specific health conditions dummy variables[14] present in all waves of data together with dummy variables derived from the general SAH measure (excellent, fair, poor/very poor health leaving good/very good health as baseline)[15] using pooled Ordinary Least Square (OLS). We

[12] Additional and updated information on the SF-36 and its related literature are available on the SF-36 community web page (http://www.sf-36.org).

[13] As specified in Brazier, Roberts and Deverill (2002), to build the SF-6D (where 6D stands for 6 dimensions) they have excluded general health items and collapsed the two dimensions of role limitations due to physical and emotional problems into a single role limitations dimension.

[14] These cover problems related to arms, legs or hands, sight, hearing, skin conditions, chest/breathing, heart/blood pressure, stomach or digestion, diabetes, anxiety/depression, alcohol or drug use, epilepsy, migraine.

[15] Due to a change in wording and response categories in the SAH question at wave 9, we collapse the original five category self-assessed variable (SAH) to a four category measure. In waves 1-8 and 10-18, respondents are asked: Compared to people of your own age, would you say your health over the last 12 months on the whole has been: excellent, good, fair, poor or very poor?, whereas in wave 9, the question and possible answers are: In general, would you say your health is: excellent, very good, good, fair, poor?. Creating a SAH variable with four health categories (excellent, good or very good, fair, poor or very poor) allows common support over the two versions of the question.

estimate separate models for men and women and use the predicted SF-6D scores from these regressions as our measure of health in our main empirical specifications.[16] In this way, we obtain a cardinal measure of health for all individuals in our survey, defined on a continuous scale from 0 to 1 which is used as our empirical proxy of health capital. As the measure draws from a wide range of health domains we feel that it better approximates the notion of health capital as originally defined in the Grossman model.

5.4 Socio-economic and demographic variables

5.4.1 Time-varying regressors

Our dynamic models of health and smoking are estimated separately for men and women and are conditioned on age and age squared; household characteristics (being married or cohabiting, household size and the number of cohabiting children); labour status (employed, self-employed, unemployed, retired, long-term sick/disability status and other employment[17]) with employed as an employee as the baseline and household income (equivalised annual log-household income.). In order to account for the effects of health shocks on both smoking preferences and health status, we include a dummy variable defining the presence of health shocks/accidents that led to hospitalisation in the previous year.

5.4.2 Time-invariant regressors

In addition to the set of regressors outlined above, we include a set of time-invariant variables for highest attained educational qualification (in descending order: degree or higher degree, HND or A-level, O-level or CSE, versus no qualification), ethnicity (categorised as white versus non-white) and a set of region of residence dummy variables.[18]

In addition, a vector of year dummies is included in all models to account for aggregate health shocks, time-varying reporting changes, trends in smoking

[16]We have also employed alternative specifications to compute predicted SF-6D values for all individuals in the sample such as linear fixed effects models. We have also estimated versions of these models (pooled OLS and linear fixed effects models) with lagged values of all regressors (health variables) to ease potential problems related to endogeneity. As predicted SF-6D scores and results from the main dynamic empirical models of health capital and smoking do not appear to differ across these specifications; we use results from the simple pooled OLS models to maximise the number of observations.

[17]Other employment consists of: looking after the family, maternity leave, government training, student or other jobs.

[18]The set of regional dummy variables contains little variation across the waves and accordingly these are categorised as the region in which a respondent was observed to reside the longest. The regions cover England (South East, South West, London, Midlands, Yorkshire, North West, North East), Scotland, Wales and Northern Ireland. The South East is taken as the baseline.

incidence and prevalence and any effects of ageing not captured by the age variables.

Descriptive statistics for the set of explanatory variables are presented in Table 1 separately for men and women. To save space these are presented for the sub-sample of individuals used in the estimation of the smoking models. These models contain a larger number of respondents than the corresponding health equation. On average men smoke more cigarettes per day than women (2.234 versus 2.123 half-packs respectively) and report better health status (0.810 versus 0.785). The two samples are of similar mean age. Men are more likely to be married or cohabiting, and more likely to be employed or self-employed than women. Women are more likely to be catergorised as other employed or retired than men. A larger proportion of men have a degree or higher degree, or HND/A-level qualification compared to the sample of women. Men report a higher household income than women and more health shocks and/or accidents resulting in hospitalisation. A larger proportion of women in the sample report white ethnic origin.

6 Results and discussion

6.1 Regression models

Table 2 for men and Table 3 for women summarise our estimation results separately for models for health and smoking. The first column presents OLS estimates of models (9) and (10) above, applying robust standard errors to capture general forms of heteroscedasticity. OLS estimation of dynamic panel data models are biased upwards ((Nickell, 1981)), however, the estimated coefficients on the lags of the dependent variable exhibit clear and strong gradients for both health and smoking. While our estimates represent structural effects of composite parameters, it is worth noting that state dependence in smoking and health outcomes have been observed elsewhere (for example, Baltagi and Levin (1986); Christelis and Sanz-de-Galdeano (2009); Contoyannis et al. (2004)). The second column presents corresponding estimates from within (fixed effects) estimation. These are biased downwards and while showing a gradient across the lagged terms, the effects are smaller than corresponding OLS estimates and are less significant statistically.

The third column presents results of system GMM estimation of models (9) and (10). Given the moving-average of order 3 in the errors, instruments are constructed from observations of the dependent variable from period $t-5$ and before for the model in first-differenced form. Estimated coefficients are expected to fall between OLS and within estimation. Estimation was performed by varying the maximum number of lags of the dependent variable from which to define instruments, and results reported for the specification that produced the most credible estimates. This decision was based on the Sargan test for over-identification, tests for serial correlation and judgement

on the resulting estimates. For example, coefficients closer to within than OLS estimation may be indicative of weak instruments. For men, instruments were constructed from observations between 5 and 6 lagged periods for health and 5 and 7 lagged periods for smoking. For women these were between 5 and 8 lagged periods for both health and smoking models. Given the long lag period required to construct instruments, it is not surprising that the resulting instruments are weak leading to estimated coefficients on the lags of the dependent variable lying outside the range of OLS and within estimation. This holds for the majority of estimates across both health and smoking models, for both men and women. These estimates are not reliable.

Tests for serial correlation in first-differenced form reveal, in general, correlation of order 1 (reported in Tables 2 and 3). This corresponds to a lack of moving average terms in the levels error structure of (9) and (10). Imposing the restriction that $\rho_{\epsilon 1}, \rho_{\epsilon 2}, \rho_{\epsilon 3} = 0$ in (9), and $\rho_{\varepsilon 1}, \rho_{\varepsilon 2}, \rho_{\varepsilon 3} = 0$ in (10) the respective error terms can simply be represented as ϵ_{it} and ε_{it}. These restrictions free up instruments from periods closer to lags of the dependent variable, $S_{it-1}, \ldots, S_{it-4}$, and $H_{it-1}, \ldots, H_{it-4}$. More recent lags are likely to have greater predictive power and hence greater relevance as instruments than those constructed from periods further away from the lagged terms.[19] Column 4 presents GMM system results assuming a lack of a moving average process (MA(0)) in the level's error. For models of both health and smoking, and for men and women, this specification is supported by tests for first-order serial correlation in first-differenced form and Sargan tests for over-identification. Parameter estimates on the lags of the dependent variable lie between OLS and within estimates and generally are closer to the former. For women, all lagged terms are significant at conventional (5%) levels; for men, the first three lags are significant. These results indicate strong and enduring persistence in the evolution of both health capital and smoking consumption with direct effects on current health and smoking observed for up to three and four lagged periods for men and women respectively.

All models contain contemporaneous values of the set of exogenous regressors X_{it} and their corresponding lags: $X_{it-1}, X_{it-2}, X_{it-3}$, together with the set of time-invariant regressors, W_i, and a vector of year dummy variables. Many of the regressors, X_{it}, display little variation over time (for example, marital status, employment status) and accordingly suffer from collinearity. Few of the lagged terms are significant in the model presented in column 4.[20] To simplify the interpretation of the effects of these regressors, we restrict the coefficients on the lagged terms to be zero. This results in estimates presented in column 5. Again, the models pass relevant specification tests and the coefficients on the lagged dependent terms do not change substantively from those of column 4.

[19] $S_{it-2}, \ldots, S_{it1}$ are potential instruments for $\Delta S_{it-1}, \ldots, \Delta S_{it-4}$. Similarly, for $\Delta H_{it-1}, \ldots, \Delta H_{it-4}$.

[20] These results are available on request.

Table 4 presents estimated coefficients for the set of covariates X_{it} and W_i.[21] For men, larger household income is associated with increased health and other employment is associated with decreased health (at the 6% significance level). Interestingly individuals who reported an accident leading to a hospitalization report better health than those not reporting an accident. Living in the South West (compared to the South East) is associated with decreased health status (at the 10% level). For women, lower health status is reported for those living in Yorkshire (at 10%), the North West, the North East, Scotland, Wales and Northern Ireland (at the 6% significance level) compared to the South East. Again an accident in the previous twelve months is associated with reporting higher health status. For male ever smokers, being unemployed (at the 5% level) or other employment is associated with lower consumption of cigarettes (at the 10% level) as is being married or cohabiting (at the 10% level). Men living in Scotland, Wales and Northern Ireland report higher consumption than men living in the South East (all at 5% significance). Being married or cohabiting is also associated with lower levels of smoking for women, while a greater number of children is associated with higher smoking prevalence. Being self-employed is also associated with a higher consumption of cigarettes (at the 10% level). For women there is a clear educational gradient with higher educated individuals smoking less than lower educated individuals (baseline is no qualifications). These effects are highly significant. Women belonging to non-white ethnic groups smoke less than their white couterparts. There are also clear regional effects, with women living in the North West, North East, Scotland, and Wales all reporting higher levels of cigarette consumption than those living in the South East of England.

6.2 Characteristic roots

The observed evolution of an individual's stock of health (or smoking) will depend in part on changes in exogenous variables and in part on the intrinsic dynamics inherent in their optimal lifetime trajectory. That trajectory is characterized by what is sometimes referred to as path dependence, conditional on the values of the exogenous variables. Whether any series of observations on H (or S) are on the same trajectory depends on the frequency with which the exogenous variables cause the trajectory to shift. The trend along this trajectory is in most cases non-linear. The presence of the exogenous explanatory variables in the equation means that we do a better job of estimating the characteristics of the lifetime trajectory, since they will control for shifts in the trajectory which are due to changes in the value of the exogenous variables. At the same time controlling for the natural tendency of the variable in question to evolve over time means that our estimation of the coefficients on the exogenous variables will be more efficient.

[21] The year dummies, which are not reported, indicate a decreasing trend in smoking across the waves for men but not for women and no discernible trends for health.

The intrinsic dynamics of H and S are characterised by four roots. For men the health equation has roots: $0.787, -0.363$, and -0.026 ± 0.417. For men smoking the roots are: $0.809, -0.225$, and -0.04 ± 0.294. Since the roots are highly non-linear combinations of the coefficients we cannot test the hypothesis that the corresponding roots match across the equations, as should be the case given our theoretical structure. Instead we comment on the general pattern of the results.

Both equations have two real and two complex roots. The dominant real roots which drive the long term behaviour of the variables for health and smoking are both positive and less than one in absolute value, and are, numerically very similar. The second real roots are both negative, both less than 1 in absolute value and of similar magnitude across the equations. While negative roots are unusual in economic models, they can arise in empirical applications simply as a consequence of the evolution of the variable in question between observation points. The final two roots are, in each equation, complex conjugate pairs, implying a cyclical element to the trajectory. The modulus of the complex roots in the health equation is 0.418 and in the smoking equation 0.297. The general pattern of the trajectories across the two equations are, therefore, fairly similar, particularly with respect to the dominant root.

For the health equation for women, the roots are: $0.791, -0.409$, and -0.014 ± 0.412, and for smoking: $0.812, -0.349$, and -0.002 ± 0.356. Again the dominant roots are very similar across the pair of equations (and very similar to the dominant roots in the male equations, which is not required by the model). Again the second real root is negative in each equation, less than 1 in absolute value and reasonably similar across the equations, and the final pairs are complex conjugates. In the health equation for women the modulus of the complex roots is 0.412 and in the smoking equation; 0.356. Interestingly, for both men and women the complex roots all imply cycles of periodicity of roughly four years. It seems likely that this is an artefact of the data.

7 Discussion

A key insight of the Grossman model is the recognition that health related behaviour has the characteristics of an investment in human capital, and accordingly a great many health-related decisions can be seen as elements of an individual's inter-temporal optimization problem. While the theoretical foundations of the Grossman model are well understood, its empirical implementation has often proved difficult. A reliance on cross-sectional or longitudinal data with small T clearly poses serious challenges to estimation of the dynamic relationships fundamental to the model. In an attempt to circumvent these issues, one strand of the empirical literature has followed the approach adopted by Wagstaff (1986).[22] This involves using the theoretical model to

[22]In turn Wagstaff (1986) was following suggestions made by Grossman (1972).

find an equation representing an individual's optimal choice of health investment as a function of the levels of exogenous variables and the rates of change in certain of the exogenous variables and assuming that the dynamic element can be assigned to the regression residual term. This has required empirical researchers to assume that changes to the observable exogenous variables, which shift the individual's optimal trajectory, were larger than the intrinsic period-to-period dynamic progression along those trajectories, and that the exogenous variables changed, discretely, with sufficient frequency that shifts of the trajectories would more than dominate movements along the trajectories through the span of the available data. This is not an entirely satisfactory solution and does not fully respect the way in which dynamics appear in the theoretical model.

Grossman (2000) discusses an empirical version of the model that would be appropriate for implementation with individual-level longitudinal data. He specifies a second order difference equation in health capital, H_t, as a function of H_{t-1} and H_{t+1} and relevant prices, together with health investments in period I_{t-1} written as a function of H_{t-1} and H_{t+1} and relevant prices but not of other periods values of I.[23] The second-order difference equations can be solved to express H_t or I_{t-1} as functions of current, past and future values of all the exogenous variables. At a minimum three waves of data are required to identify the model, with additional waves needed should instruments be constructed from past periods.

It is worth noting that the structural forms which Grossman proposes are very similar to forms which have regularly been estimated in the related literature on RA. This is not surprising since the RA model is basically a health investment model in which the most interesting cases are those of goods that are, over the long run, bad for your health and for which the appetite grows. The empirical RA literature, however, has most commonly made use of market level rather than individual level data, although that has changed more recently (for example, Labeaga (1999)).[24] Arguably less use has been made of individual level data for estimation of pure Grossman type models.

This paper makes use of a mature British panel data set containing up to eighteen waves of data on any given survey respondent. This allows us to robustly estimate an empirical specification containing structural dynamic elements derived directly from a theoretical model that combines Grossman's health investment concept and the B-M RA model. In turn, this allows greater understanding of the inherent dynamics of the model, but also the impact of exogenous variables assumed to shift an individual's optimal trajectory.

It is important to note here that we are not dealing with macro data, or even market level data, in which variables are likely to be trended, either exogenously or through the presence of a unit root. While our dominant roots

[23]These relate to equations (55) and (56) in Grossman (2000).

[24]Empirical results derived from individual level data do not contain unit roots and are in general consistent with theoretical RA dynamics

are reasonably large, and although lack of standard errors means that we cannot formally test them against unity, there does not appear to be a non-stationarity problem in our data set. The lagged dependent variables in this type of equation are representations of intrinsic features of an individual's optimal trajectory.

Conditional on the intrinsic dynamics in the health and smoking equations, few of the coefficients on the exogenous explanatory variables are statistically significant. For men higher household income is associated with being healthier while for women health appears independent of income. The effect for men is small, with a 1 unit increase in log equivalised income (calculated at the mean this relates to a substantial increase in income from 10,360 to 36,315) leading to a 0.5% point increase in SF-6D values (on a scale of 0 to 1). For women health status appears independent of income. These small impacts of income on health appear to corroborate findings elsewhere on the relationship between income and health. Using the fall of the Berlin Wall as a natural experiment to study the impact of rapid increases in income for East Germans following reunification, Frijters et al. (2005), found a similarly small impact on men (a 1 log point increase in income led to a 0.083 increase in health satisfaction measured on a latent scale corresponding to an observed ordinal scale of 0 to 10). They reported no effect for East German women.[25] Contoyannis et al. (2004) and Meer (2003) also report small gradients in the income-health relationship. Using a panel of elderly Americans, Adams et al. (2003), develop tests of no direct causal link between socio-economic status and health conditional on initial past health status, and do not find evidence to the contray when considering mortality and incidence of most sudden onset of health conditions (accidents and some acute conditions), but some association with the incidence of gradual onset of health conditions (mental health, and some degenerative and chronic conditions). There main measure of socio-economic status is wealth, which appears to have far greater association with health than current income (as used in our study).

Interestingly, highest educational qualifications while generally displaying the expected gradients for health (positive) for women, is not statistically significant. While this is contrary to theoretical predictions from the pure Grossman model (Grossman, 1972), this finding is supported by those of Adams et al. (2003) who also conclude that education, conditional on socio-economic status (wealth), is not systematically associated with health. Education does, however, display a more pronounced gradient for smoking with effects greater (and statistically significant) for women compared to men. For the sample of female ever-smokers, having an O-level or GCSE qualification is associated with 0.65 less half packs smoked daily (equivalent to three and a quarter cigarettes) and attaining a degree or higher degree compared to HND/A-Level

[25]The study reported significant effects of income on health satisfaction for West Germans for both men and women, although these effects were smaller than observed for men in East Germany.

qualification corresponds, on average, to a decrease in 0.2 packs of cigarettes a day (equating to around a single cigarette daily). Being married or cohabiting is associated with smoking less for both men and women; a results observed elsewhere (Linström, 2010). Non-white ethnicity is significantly associated with reduced daily smoking for women. In general, for both men and women non-white ethnic groups have a lower prevalence of smoking than white ethnic groups, although this masks important variation across minority groups which is more pronounced for women than for men. For example, Black-Caribbean and Other South Asian women have a far greater prevalence of regular smoking (at levels slightly lower than white ethnic groups) compared to Bangladeshi, Indian, Pakistani , Chinese and other Black minority ethnic groups (Mellward and Karlson, 2011). There is evidence of geographical variation, particularly for women where areas associated with decreased health (compared to the baseline of South East of England) are also associated with increased smoking. While this is likely to reflect variation in area deprivation, there is, however, the possibility that geographic dummies are picking up the effect of differences across regions in the distributions of the other explanatory variables.

One result that may appear surprising is the general lack of significance of the terms for age. We typically expect age to play a significant role, at least in an equation for health. In other empirical work, especially studies relying on cross-section data and to a slightly lesser extent very short panel data studies, age has acted as a proxy for the stage of an individual along her lifetime trajectory. Given the tendency of that trajectory to non-linearity, age works best as a proxy when entered as a polynomial. In our theoretical model, though, age enters as a determinant of the rate of depreciation of health capital, with depreciation increasing at older ages. This increase in the rate of depreciation will tend to cause the individual's stock of health capital to decline faster in later years than in earlier years, but the individual may respond to this increased rate of depreciation by increasing her investment in health and slowing the rate of decline of H. This is part of the process by which the optimizing individual contrives to follow the optimal lifetime trajectory for health capital. Given that we have incorporated the shape of the trajectory directly into the estimation by running fourth order difference equations, there remains little role for age to play.

Of the remaining variables, the lack of significance may well be a consequence of a lack of within-individual variation. Alternatively it may indicate that these variables do not have a role in shifting the optimal trajectory for an individual. If these are variables which do not change often and which have relatively small impacts on the position of the individual's optimal trajectory, it is not surprising that, conditional on the intrinsic dynamics, they do not appear important.

Despite the econometric challenges associated with our approach, we argue that further advances in our understanding of the essence of individual health related decisions, given that those decisions are so clearly inter-related, are

most likely to follow from the application of micro-econometric techniques such as those used here to detailed structural modelling. The increased availability of long panel data sets containing a range of health and general economic variables opens a rich frontier for developments in health economics with a firm theoretical foundation. Ultimately this is the type of approach needed if we are properly to test the applicability of the human capital model to health-related behaviours.

	Men NT = 18407				Women NT = 21915			
Variables	Mean	SD	Min	Max	Mean	SD	Min	Max
Cigarettes (1/2 packs)	2.234	2.096	0	9	2.123	1.862	0	9
Health	0.810	0.070	0.462	0.908	0.785	0.080	0.467	0.899
Age	45.25	15.96	19	98	45.26	15.98	19	96
Married/Co-habiting	0.710	0.454	0	1	0.632	0.482	0	1
Household Size	2.866	1.398	1	16	2.806	1.333	1	16
Number of Children	0.565	0.978	0	7	0.655	0.995	0	8
Employed	0.568	0.495	0	1	0.517	0.500	0	1
Unemployed	0.058	0.223	0	1	0.030	0.171	0	1
Self Employed	0.115	0.319	0	1	0.036	0.187	0	1
Retired	0.156	0.363	0	1	0.173	0.378	0	1
Employment other	0.089	0.284	0	1	0.227	0.419	0	1
Long-term sick	0.014	0.119	0	1	0.016	0.127	0	1
Log Household Income	9.636	0.646	-0.350	12.914	9.552	0.637	-0.174	13.505
Accidents	0.119	0.284	0	1	0.087	0.283	0	1
Degree/Higher degree	0.104	0.443	0	1	0.095	0.293	0	1
HND/A-Level	0.268	0.443	0	1	0.191	0.393	0	1
O-Level/ CSE	0.304	0.460	0	1	0.344	0.475	0	1
No Qualifications	0.324	0.468	0	1	0.371	0.483	0	1
White	0.970	0.170	0	1	0.987	0.111	0	1
Non-White	0.030	0.170	0	1	0.013	0.111	0	1
South East	0.149	0.356	0	1	0.148	0.355	0	1
South West	0.068	0.253	0	1	0.062	0.240	0	1
London	0.074	0.261	0	1	0.067	0.250	0	1
Midlands	0.193	0.395	0	1	0.169	0.375	0	1
Yorkshire	0.082	0.275	0	1	0.081	0.273	0	1
North West	0.091	0.288	0	1	0.094	0.292	0	1
North East	0.045	0.207	0	1	0.054	0.225	0	1
Scotland	0.133	0.340	0	1	0.149	0.356	0	1
Wales	0.110	0.313	0	1	0.111	0.314	0	1
Northern Ireland	0.054	0.227	0	1	0.066	0.248	0	1

Table 1: Descriptive statistics. Sample based on Smoking equations (NT = 18407 for men and NT = 21915 for women, except for the health variable (NT = 17918 for men and NT = 21429 for women

Health	OLS	Within	System GMM		
			MA(3)	MA(0)	MA(0)
	(1)	(2)	(3)	(4)	(5)
H_{it-1}	0.412 (0.000)	0.122 (0.000)	0.510 (0.014)	0.367 (0.000)	0.372 (0.000)
H_{it-2}	0.177 (0.000)	-0.001 (0.918)	0.340 (0.093)	0.133 (0.000)	0.132 (0.000)
H_{it-3}	0.140 (0.000)	-0.007 (0.573)	0.226 (0.178)	0.087 (0.002)	0.088 (0.001)
H_{it-4}	0.104 (0.000)	-0.040 (0.000)	-0.062 (0.104)	0.043 (0.220)	0.051 (0.128)
X_{it}	Y	Y	Y	Y	Y
X_{it-1}	Y	Y	Y	Y	N
X_{it-2}	Y	Y	Y	Y	N
X_{it-3}	Y	Y	Y	Y	N
W_i	Y	N	Y	Y	Y
Years	Y	Y	Y	Y	Y
NT (N)	14635 (2315)	14635 (2315)	14635 (2315)	14635 (2315)	14635 (2315)
Sargan test			22.7[31] (0.858)	52.0[45] (0.220)	51.4[45] (0.238)
Serial Corr:					
Order (1)			-2.22 (0.026)	-20.2 (0.000)	-20.42 (0.000)
Order (2)			-0.29 (0.771)	0.30 (0.763)	0.32 (0.749)
Order (3)			-0.67 (0.502)		
Order (4)			1.25 (0.210)		
Order (5)			1.20 (0.229)		
Lags:			(5 6)	(2 4)	(2 4)

Smoking	OLS	Within	System GMM		
			MA(3)	MA(0)	MA(0)
S_{it-1}	0.540 (0.000)	0.296 (0.000)	0.360 (0.087)	0.509 (0.000)	0.504 (0.000)
S_{it-2}	0.181 (0.000)	0.063 (0.000)	0.461 (0.003)	0.143 (0.000)	0.142 (0.000)
S_{it-3}	0.093 (0.000)	0.011 (0.327)	-0.046 (0.775)	0.067 (0.000)	0.066 (0.000)
S_{it-4}	0.066 (0.000)	-0.023 (0.024)	0.034 (0.405)	0.016 (0.231)	0.016 (0.250)
X_{it}	Y	Y	Y	Y	Y
X_{it-1}	Y	Y	Y	Y	N
X_{it-2}	Y	Y	Y	Y	N
X_{it-3}	Y	Y	Y	Y	N
W_i	Y	N	Y	Y	Y
Years	Y	Y	Y	Y	Y
NT (N)	18407 (2864)	18407 (2864)	18407 (2864)	18407 (2864)	18407 (2864)
Sargan test			34.7[45] (0.867)	76.7[62] (0.099)	74.7[62] (0.130)
Serial Corr:					
Order (1)			-2.18 (0.029)	-20.78 (0.000)	-20.74 (0.000)
Order (2)			-2.94 (0.003)	1.07 (0.283)	0.92 (0.359)
Order (3)			1.29 (0.198)		
Order (4)			-0.29 (0.771)		
Order (5)			0.64 (0.519)		
Lags:			(5 7)	(2 5)	(2 5)

Table 2: Men: Fourth-order single equation estimates. Coefficient estimates and p-values in parentheses. The Sargan test reports the statistic, degrees of freedom [ ] and associated p-value ( ). Tests for serial correlation in first-differenced errors report the test statistic and p-value ( ). Lags reports the lag structure used to construct instruments for the model in first-differened form. Two-step robust standard errors are used (Windmeijer, 2005).

Health	OLS	Within	System GMM		
			MA(3)	MA(0)	MA(0)
	(1)	(2)	(3)	(4)	(5)
H_{it-1}	0.407 (0.000)	0.119 (0.000)	0.403 (0.009)	0.341 (0.000)	0.355 (0.000)
H_{it-2}	0.200 (0.000)	0.015 (0.144)	0.205 (0.160)	0.158 (0.000)	0.164 (0.000)
H_{it-3}	0.131 (0.000)	-0.015 (0.124)	0.131 (0.335)	0.070 (0.000)	0.074 (0.000)
H_{it-4}	0.125 (0.000)	-0.021 (0.034)	0.023 (0.358)	0.051 (0.001)	0.055 (0.000)
X_{it}	Y	Y	Y	Y	Y
X_{it-1}	Y	Y	Y	Y	N
X_{it-2}	Y	Y	Y	Y	N
X_{it-3}	Y	Y	Y	Y	N
W_i	Y	N	Y	Y	Y
Years	Y	Y	Y	Y	Y
NT (N)	17674 (2701)	17674 (2701)	17674 (2701)	17674 (2701)	17674 (2701)
Sargan test			48.3[50] (0.543)	80.2[68] (0.149)	79.0[68] (0.170)
Serial Corr:					
Order (1)			-4.11 (0.000)	-25.03 (0.000)	-25.01 (0.000)
Order (2)			0.10 (0.920)	0.16 (0.873)	0.05 (0.961)
Order (3)			-0.41 (0.684)		
Order (4)			0.60 (0.552)		
Order (5)			-0.40 (0.687)		
Lags:			(5 8)	(2 6)	(2 6)
Smoking	OLS	Within	System GMM		
			MA(3)	MA(0)	MA(0)
S_{it-1}	0.520 (0.000)	0.265 (0.000)	0.171 (0.412)	0.457 (0.000)	0.458 (0.000)
S_{it-2}	0.198 (0.000)	0.075 (0.000)	0.488 (0.024)	0.159 (0.000)	0.159 (0.000)
S_{it-3}	0.098 (0.000)	0.012 (0.219)	0.077 (0.624)	0.060 (0.000)	0.060 (0.000)
S_{it-4}	0.078 (0.000)	-0.010 (0.292)	0.003 (0.931)	0.034 (0.004)	0.037 (0.001)
X_{it}	Y	Y	Y	Y	Y
X_{it-1}	Y	Y	Y	Y	N
X_{it-2}	Y	Y	Y	Y	N
X_{it-3}	Y	Y	Y	Y	N
W_i	Y	N	Y	Y	Y
Years	Y	Y	Y	Y	Y
NT (N)	21915 (3340)	21915 (3340)	21915 (3340)	21915 (3340)	21915 (3340)
Sargan test			24.4[34] (0.887)	63.4[74] (0.804)	62.0[74] (0.838)
Serial Corr:					
Order (1)			-1.95 (0.052)	-21.98 (0.000)	-22.13 (0.000)
Order (2)			-2.42 (0.016)	-0.95 (0.344)	-0.96 (0.339)
Order (3)			0.46 (0.644)		
Order (4)			0.38 (0.707)		
Order (5)			1.17 (0.240)		
Lags:			(5 8)	(2 6)	(2 6)

Table 3: Women: Fourth-order single equation estimates. Coefficient estimates and p-values in parentheses. The Sargan test reports the statistic, degrees of freedom [ ] and associated p-value ( ). Tests for serial correlation in first-differenced errors report the test statistic and p-value ( ). Lags reports the lag structure used to construct instruments for the model in first-differened form. Two-step robust standard errors are used (Windmeijer, 2005).

	Men				Women			
	Health NT = 14635		Smoking NT = 18407		Health NT = 17674		Smoking NT = 21915	
Y_{it-1}	0.372	(0.000)	0.504	(0.000)	0.355	(0.000)	0.458	(0.000)
Y_{it-2}	0.132	(0.000)	0.142	(0.000)	0.164	(0.000)	0.159	(0.000)
Y_{it-3}	0.088	(0.001)	0.066	(0.000)	0.074	(0.000)	0.060	(0.000)
Y_{it-4}	0.051	(0.128)	0.016	(0.250)	0.055	(0.000)	0.037	(0.001)
Age	-0.0005	(0.607)	-0.015	(0.518)	0.001	(0.445)	-0.033	(0.085)
Age squared	-0.00005	(0.959)	0.014	(0.494)	-0.001	(0.086)	0.005	(0.757)
Married/Co-habiting	0.004	(0.127)	-0.137	(0.078)	0.003	(0.204)	-0.121	(0.021)
Household Size	0.002	(0.126)	-0.016	(0.567)	-0.0002	(0.833)	0.014	(0.485)
Number of Children	0.0002	(0.911)	0.031	(0.532)	-0.0004	(0.843)	0.158	(0.000)
Unemployed	-0.00008	(0.981)	-0.189	(0.022)	-0.002	(0.531)	-0.032	(0.629)
Self Employed	-0.003	(0.389)	0.001	(0.986)	-0.004	(0.387)	0.162	(0.077)
Retired	-0.003	(0.499)	-0.078	(0.481)	-0.0008	(0.788)	-0.032	(0.629)
Employment other	-0.009	(0.055)	-0.161	(0.100)	0.001	(0.647)	-0.068	(0.107)
Long-term sick	-0.002	(0.739)	0.094	(0.482)	0.005	(0.340)	-0.050	(0.537)
Household Income	0.005	(0.001)	0.028	(0.337)	0.002	(0.146)	-0.002	(0.915)
Accidents	0.003	(0.027)	-0.023	(0.537)	0.004	(0.035)	0.040	(0.218)
Degree/Higher degree	0.0002	(0.984)	-0.328	(0.297)	0.011	(0.518)	-0.913	(0.006)
HND/A-Level	-0.00007	(0.994)	-0.091	(0.711)	0.010	(0.507)	-0.707	(0.016)
O-Level/ CSE	-0.002	(0.846)	-0.070	(0.778)	0.009	(0.516)	-0.650	(0.014)
Non-White	-0.006	(0.177)	-0.042	(0.727)	-0.003	(0.660)	-0.325	(0.011)
South West	0.005	(0.093)	-0.022	(0.772)	0.002	(0.406)	-0.014	(0.825)
London	0.002	(0.414)	0.067	(0.362)	0.0004	(0.904)	0.082	(0.276)
Midlands	0.0005	(0.823)	-0.031	(0.630)	-0.005	(0.117)	0.0003	(0.996)
Yorkshire	-0.002	(0.432)	-0.033	(0.610)	-0.006	(0.091)	0.073	(0.238)
North West	-0.0005	(0.817)	-0.003	(0.956)	-0.003	(0.238)	0.162	(0.014)
North East	-0.004	(0.196)	0.048	(0.549)	-0.011	(0.005)	0.168	(0.017)
Scotland	-0.002	(0.406)	0.217	(0.000)	-0.007	(0.009)	0.242	(0.000)
Wales	-0.001	(0.589)	0.179	(0.003)	-0.011	(0.004)	0.146	(0.017)
Northern Ireland	-0.003	(0.614)	0.236	(0.018)	-0.006	(0.342)	0.051	(0.630)

Table 4: Fourth-order single equation estimates. Coefficient estimates and p-values in parentheses. $Y_{it-1}, \ldots, Y_{it-4} = H_{it-1}, \ldots, H_{it-4}$ for the health equation and $S_{it-1}, \ldots, S_{it-4}$ for the smoking equation. All regressions include year dummies.

Acknowledgements

The authors gratefully acknowledge funding from the Economic and Social Research Council (ESRC) under grant reference RES-060-25-0045. We are grateful to Victor Aguirregabiria, Jonneke Bohlaar, Brenda Gannon, Eleonora Fichera, Aloysius Siow, Matt Sutton, and participants of the 8th International Health Economics Association (iHEA) Congress in Toronto, the 9th European Conference on Health Economics (ECHE) in Zurich, the 2012 Applied Health Econometrics Symposium at the University of Leeds, the 2012 Netspar Conference on Health and Inequality in the Life-cycle in Rotterdam as well as participants of seminars at Universities of Barcelona, Leeds, Newcastle, Toronto and York for discussions and helpful comments. Data from the British Household Panel Survey were supplied by the Economic and Social Research Council Data Archive. Neither the original collectors of the data nor the Archive bear any responsibility for the analysis or interpretations presented in this manuscript.

References

Adams, P., Hurd, M.D., McFadden, D., Merrill, A., Ribeiro, T. (2003). Healthy, wealthy, and wise? Tests for direct causal paths between health and socioeconomic status. *Journal of Econometrics*, 112: 3-56.

Adda, J., Lechene, V. (2013). Health selection and the effect of smoking on mortality. *The Scandinavian Journal of Economics*, 115(3): 902-931.

Adda, J., Cornaglia, F. (2006). Taxes, cigarette consumption, and smoking intensity. *The American Economic Review*, 96(4): 1013-1028.

Adda, J., Cornaglia, F. (2010). The effect of bans and taxes on passive smoking. *American Economic Journal: Applied Economics*, 2(1): 1-32.

Arcidiacono, P. (2007) Living Rationally Under the Volcano? An Empirical Analysis of Heavy Drinking and Smoking. *International Economic Review*, 48(1): 37-65.

Arellano, M., Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations, *The Review of Economic Studies* 58(2): 277-297.

Balia, S., Jones, A.M. (2008). Mortality, lifestyle and socio-economic status. *Journal of Health Economics*, 27(1); 1-26.

Baltagi, B. H. and Griffin, J. M. (2002). Rational addiction to alcohol: panel data analysis of liquor consumption. *Health Economics*, 11(6): 485-491.

Baltagi, B. H. and Levin, D. (1986). Estimating dynamic demand for cigarettes using panel data: the effects of bootlegging, taxation and advertising reconsidered. *The Review of Economics and Statistics*, 68(1): 148-155.

Becker, G.S. (1965). A Theory of the Allocation of Time. *Economic Journal*, 75(299): 493-517.

Becker, G.S., Woytinsky, W.S. (1993). *Human capital and the personal distribution of income: An analytical approach.* 1967. Institute of Public Administration, University of Michigan, Ann Arbor. Also available in: G.S. Becker. Human Capital (3rd edition). University of Chicago Press. 102-158.

Becker, G. S., Grossman, M. and Murphy, K. M. (1994). An Empirical Analysis of Cigarette Addiction. *American Economic Review*, 84(3): 396-418.

Becker, G. S. and Murphy, K. M. (1988). A Theory of Rational Addiction. *The Journal of Political Economy*, 96(4): 675-700.

Blundell, R., Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel-data models. *Journal of Econometrics*, 87(1): 115-143.

Blundell, R., Bond, S., Windmeijer, F. (2001). Estimation in dynamic panel-data models: Improving the performance of the standard GMM estimator. In *Nonstationary Panels, Cointegrating Panels and Dynamic Panels,*, Advances in Econometrics, Volumn 15, ed. Baltagi, B., Fomby, T.b., Carter Hill, R. 53-92. New York, Elsevier.

Brazier J., Roberts J., Deverill, M. (2002). The estimation of a preference-based measure of health from the SF-36. *Journal of Health Economics*, 21(2): 271-292.

Chaloupka, F. (1991). Rational Addictive Behavior and Cigarette Smoking. *Journal of Political Economy*, 99(4): 722-742.

Christelis, D., Sanz-de-Galdeano, A. (2009). *Smoking persistence across countries: an analysis using semi-parametric dynamic panel data models with sensitivity.* Discussion Paper No. 4336, Institute for the Study of Labour (IZA), July 2009.

Contoyannis, P., Jones, A.M., Rice, N. (2004). The dynamics of health in the British Household Panel Survey. *Journal of Applied Econometrics*, 19: 473-503.

Darden, M. (2012). *Smoking, Expectations, and Health: A Dynamic Stochastic Model of Lifetime Smoking Behavior.* Working Papers, Department of Economics, Tulane University. Available to download at: http://ideas.repec.org/p/tul/wpaper/1204.html

Doll, R., Peto, R., Boreham, J., Sutherland, I. (2004). Mortality in relation to smoking: 50 years' observations on male British doctors. *British Medical Journal*; doi:10.1136/bmj.38142.554479.AE

Ehrlich, I. and Chuma, H. (1990). A model of the demand for longevity and the value of life extension. *Journal of Political Economy*, 98(4): 761-782.

Ferguson, B., Lim, G.C. (2003). *Dynamic Economic Models in Discrete Time: Theory and empirical applications.* New York, Routledge.

Forster, M., (2001). The meaning of death: some simulations of a model of healthy and unhealthy consumption. *Journal of Health Economics*, 20(4); 613-638.

Frijters, P., Haisken-DeNew, J.P., Shields, M.A. (2005). The causal effect of incoem on health: Evidence from the German reunification. *The Journal of Health Economics*, 24(5): 997-1017.

Goldberg D, Williams P. (1988). *A Users Guide to the General Health Questionnaire.* Windsor: NFER-Nelson.

Grossman, M. (1972). On the Concept of Health Capital and the Demand for Health. *Journal of Political Economy*, 80(2): 223-255.

Grossman, M. (1993). The economic analysis of addictive behavior. In *Economics and the Prevention of Alcohol-Related Problems*, Hilton, M.E. Bloss, G. (eds), NIAAA Research Monograph No. 25, NIH Pub. No. 933513. Bethesda, MD: National Institute on Alcohol Abuse and Alcoholism: 91123.

Grossman, M. (2000) Chapter 7: The human capital model, *In Handbook of Health Economics, Volumn 1* Culyer, A.J., Newhouse, J.P. (eds); Elsevier, North Holland.

Grossman, M. and Chaloupka, F. J. (1998). The demand for cocaine by young adults: a rational addiction approach. *Journal of Health Economics*, 17(4); 427-474.

Grossman, M., Chaloupka, F. J. and Sirtalan, I. (1998). An empirical analysis of alcohol addiction: Results from the monitoring the future panels. *Economic Inquiry*, 36(1): 39-48.

Gruber, J., Koszegi, B. (2001). Is addiction rational? Theory and evidence. *Quarterly Journal of Economics*, 116(4): 1261-1303.

Hausman, J., Taylor, W. (1981). Panel data and unobserved individual effects. *Econometrica*, 49(6): 1377-1398.

Jones, A. M. and Labeaga, J. M. (2003). Individual heterogeneity and censoring in panel data estimates of tobacco expenditure. *Journal of Applied Econometrics*, 18(2): 157-177.

Keeler, T. E., Hu, T.-W., Barnett, P. G. and Manning, W. G. (1993). Taxation, regulation, and addiction: A demand function for cigarettes based on time-series evidence. *Journal of Health Economics*, 12(1): 1-18.

Labeaga, J. M. (1999). A Double-Hurdle Rational Addiction Model with Heterogeneity: Estimating the Demand for Tobacco. *Journal of Econometrics*, 93(1): 49-72.

Leonard, D., Long, N.V. (1992). Optimal Control Theory and Static Optimization in Economics, Chapter 10, Section 10.1 "Problems with two-state variables": pp307-310. Cambridge University Press.

Lindström, M. (2010). Social capital, economics conditions, marital status and daily smoking: A population-based study. *Pulic Health*, 124(2): 71-77.

Lynn P. (ed.) (2006). *Quality Profile: British Household Panel Survey Version 2.0: Waves 1 to 13: 1991-2003.* Colchester, Institute for Social and Economic Research, University of Essex.

Meer, J,. Miller, D., Rosen, H. (2003). Exploring the health-wealth nexus. *Journal of Health Economics*, 22: 713-730.

Mellward, D., Karlson, S. (2011). *Tobacco use among minority ethnic populations and cessation interventions.* A Race Equality Foundation Briefing Paper, London.

Mincer, J. (1974). *Schooling, Experience and Earnings.* Columbia University Press for the National Bureau of Economic Research, New York.

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica*, 49(6): 1417-1426.

Pitchford, J.D. (1977). Two state variable problems. In *Applications of Control Theory to Economic Analysis*, Pictchford, J.D., Turnovsky, S.J. (eds), North-Holland, Amsterdam: 127154

Saffer, H. and Chaloupka, F. (1999). The demand for illicit drugs. *Economic Inquiry*, 37(3): 401-411.

Wagstaff, A. (1986). The demands for health: Some new empirical evidence. *Journal of Health Economics*, 5: 195-233.

Ware, J.E., Snow, K.K., Kolinski, M., Gandeck, B. (1993). *SF-36 Health survey manual and interpretation guide.* The Health Institute, New England Medical Centre, Boston, MA.

Waters, T. M. and Sloan, F. A. (1995). Why do people drink? Tests of the rational addiction model. *Applied Economics*, 27(8); 727-736.

Windmeijer, F. (2005). A finite sample correction for the variance of linear efficient two-step GMM estimators. *Journal of Econometrics*, 126: 25-52.

8 Appendix A

Consider the four by four system of equations for H, S, A and M and where X is exogenous.

$$\begin{aligned} H_t &= \vartheta_0 + \vartheta_1 H_{t-1} + \vartheta_2 S_{t-1} + \vartheta_3 A_{t-1} + \vartheta_4 M_{t-1} + \vartheta_5 X_t + \eta_t^h \\ S_t &= \alpha_0 + \alpha_1 H_{t-1} + \alpha_2 S_{t-1} + \alpha_3 A_{t-1} + \alpha_4 M_{t-1} + \alpha_5 X_t + \eta_t^s \\ A_t &= \gamma_0 + \gamma_1 H_{t-1} + \gamma_2 S_{t-1} + \gamma_3 A_{t-1} + \gamma_4 M_{t-1} + \gamma_5 X_t + \eta_t^a \\ M_t &= \omega_0 + \omega_1 H_{t-1} + \omega_2 S_{t-1} + \omega_3 A_{t-1} + \omega_4 M_{t-1} + \omega_5 X_t + \eta_t^m \end{aligned} \quad (13)$$

Using the lag operator, such that $LH_t = H_{t-1}$ etc., and grouping yields:

$$\begin{pmatrix} (1-\vartheta_1 L) & -\vartheta_2 L & -\vartheta_3 L & -\vartheta_4 L \\ -\alpha_1 L & (1-\alpha_2 L) & -\alpha_3 L & -\alpha_4 L \\ -\gamma_1 L & -\gamma_2 L & (1-\gamma_3 L) & -\gamma_4 L \\ -\omega_1 L & -\omega_2 L & \omega_3 L & (1-\omega_4 L) \end{pmatrix} \begin{pmatrix} H_t \\ S_t \\ A_t \\ M_t \end{pmatrix} = \begin{pmatrix} \vartheta_0 + \vartheta_5 X_t + \eta_t^H \\ \alpha_0 + \alpha_5 X_t + \eta_t^S \\ \gamma_0 + \gamma_5 X_t + \eta_t^A \\ \omega_0 + \omega_5 X_t + \eta_t^M \end{pmatrix}$$

$$\begin{pmatrix} H_t \\ S_t \\ A_t \\ M_t \end{pmatrix} = \begin{pmatrix} (1-\vartheta_1 L) & -\vartheta_2 L & -\vartheta_3 L & -\vartheta_4 L \\ -\alpha_1 L & (1-\alpha_2 L) & -\alpha_3 L & -\alpha_4 L \\ -\gamma_1 L & -\gamma_2 L & (1-\gamma_3 L) & -\gamma_4 L \\ -\omega_1 L & -\omega_2 L & \omega_3 L & (1-\omega_4 L) \end{pmatrix}^{-1} \begin{pmatrix} \vartheta_0 + \vartheta_5 X_t + \eta_t^H \\ \alpha_0 + \alpha_5 X_t + \eta_t^S \\ \gamma_0 + \gamma_5 X_t + \eta_t^A \\ \omega_0 + \omega_5 X_t + \eta_t^M \end{pmatrix}$$

Solving for the inverse of the matrix in the expression above requires us to find the determinant (of the fourth order). This is given as:

$$\begin{aligned} &-1 + [\vartheta_1 + \alpha_2 - \gamma_3 + \omega_4] L \\ &+ [\vartheta_2\alpha_1 - \vartheta_1\alpha_2 - \vartheta_3\gamma_1 - \alpha_3\gamma_2 + \vartheta_1\gamma_3 + \alpha_2\gamma_3 + \vartheta_4\omega_1 + \alpha_4\omega_2 - \gamma_4\omega_3 - \vartheta_1\omega_4 \\ &-\alpha_2\omega_4 + \gamma_3\omega_4] L^2 \\ &+ [\vartheta_3\alpha_2\gamma_1 - \vartheta_2\alpha_3\gamma_1 - \vartheta_3\alpha_1\gamma_2 + \vartheta_1\alpha_3\gamma_2 + \vartheta_2\alpha_1\gamma_3 - \vartheta_1\alpha_2\gamma_3 - \vartheta_4\alpha_2\omega_1 + \vartheta_2\alpha_4\omega_1 \\ &+\vartheta_4\gamma_3\omega_1 - \vartheta_3\gamma_4\omega_1 + \vartheta_4\alpha_1\omega_2 - \vartheta_1\alpha_4\omega_2 + \alpha_4\gamma_3\omega_2 - \alpha_3\gamma_4\omega_2 - \vartheta_4\gamma_1\omega_3 - \alpha_4\gamma_2\omega_3 \\ &+\vartheta_1\gamma_4\omega_3 + \alpha_2\gamma_4\omega_3 - \vartheta_2\alpha_1\omega_4 + +\vartheta_1\alpha_2\omega_4 + \vartheta_3\gamma_1\omega_4 + \alpha_3\gamma_2\omega_4 - \vartheta_1\gamma_3\omega_4 - \alpha_2\gamma_3\omega_4] L^3 \\ &+ [\vartheta_4\alpha_3\gamma_2\omega_1 - \vartheta_3\alpha_4\gamma_2\omega_1 - \vartheta_4\alpha_2\gamma_3\omega_1 + \vartheta_2\alpha_4\gamma_3\omega_1 + \vartheta_3\alpha_2\gamma_4\omega_1 - \vartheta_2\alpha_3\gamma_4\omega_1 \\ &-\vartheta_4\alpha_3\gamma_1\omega_2 + \vartheta_3\alpha_4\gamma_1\omega_2 + \vartheta_4\alpha_4\gamma_1\omega_2 + \vartheta_4\alpha_1\gamma_3\omega_2 - \vartheta_1\alpha_4\gamma_3\omega_2 - \vartheta_3\alpha_1\gamma_4\omega_2 \\ &+\vartheta_1\alpha_3\gamma_4\omega_2 + \vartheta_4\alpha_2\gamma_1\omega_3 - \vartheta_2\alpha_4\gamma_1\omega_3 + \vartheta_4\alpha_1\gamma_2\omega_3 + \vartheta_1\alpha_4\gamma_2\omega_3 + \vartheta_2\alpha_1\gamma_4\omega_3 \\ &-\vartheta_1\alpha_2\gamma_4\omega_3 - \vartheta_3\alpha_2\gamma_1\omega_4 + \vartheta_2\alpha_3\gamma_1\omega_4 + \vartheta_3\alpha_1\gamma_2\omega_4 - \vartheta_1\alpha_3\gamma_2\omega_4 - \vartheta_2\alpha_1\gamma_3\omega_4 \\ &+\vartheta_1\alpha_2\gamma_3\omega_4] L^4 \end{aligned}$$

When applied to each of the variables, H_t, S_t, A_t or M_t on the left hand side of the expression being reduced yields a fourth order difference equation in that variable.

The adjoint matrix has the following form:

Row 1:

$$
\begin{array}{l}
-1+\left[\alpha_2-\gamma_3+\omega_4\right]L+\left[\alpha_2\gamma_3+\alpha_4\omega_2-\alpha_3\gamma_2-\gamma_4\omega_3-\alpha_2\omega_4\right]L^2 \\
+\left[\alpha_4\gamma_3\omega_2-\alpha_3\gamma_4\omega_2-\alpha_4\gamma_2\omega_3+\alpha_2\gamma_4\omega_3+\alpha_3\gamma_2\omega_4-\alpha_2L\gamma_3L\omega_4L\right]L^3, \\
-\vartheta_2L+\left[\vartheta_3\gamma_2-\vartheta_2\gamma_3-\vartheta_4\omega_2+\vartheta_2\omega_4\right]L^2+\left[\vartheta_3\gamma_4\omega_2-\vartheta_4\gamma_3\omega_2+\vartheta_4\gamma_2\omega_3\right. \\
\left.-\vartheta_2\gamma_4\omega_3+\vartheta_3\gamma_2\omega_4+\vartheta_2\gamma_3\omega_4\right]L^3, \\
\vartheta_3L+\left[\vartheta_2\alpha_3-\vartheta_3\alpha_2+\vartheta_4\omega_3-\vartheta_1\omega_4\right]L^2+\left[\vartheta_4\alpha_3\omega_2-\vartheta_3\alpha_4\omega_2-\vartheta_4\alpha_2\omega_3\right. \\
\left.+\vartheta_2\alpha_4\omega_3+\vartheta_3\alpha_2\omega_4-\vartheta_2\alpha_3\omega_4\right]L^3, \\
-\vartheta_4L+\left[\vartheta_3\gamma_4+\vartheta_4\alpha_2-\vartheta_2\alpha_4-\vartheta_4\gamma_3\right]L^2+\left[\vartheta_3\alpha_4\gamma_2-\vartheta_4\alpha_3\omega_2+\vartheta_4\alpha_2\gamma_3\right. \\
\left.-\vartheta_2\alpha_4\gamma_3-\vartheta_3\alpha_2\omega_4+\vartheta_2\alpha_3\gamma_4\right]L^3.
\end{array}
$$

Row 2:

$$
\begin{array}{l}
-\alpha_1L+\left[\alpha_3\gamma_1-\alpha_1\gamma_3-\alpha_4\omega_1+\alpha_1\omega_4\right]L^2 \\
+\left[\alpha_3\gamma_4\omega_1+\alpha_4\gamma_1\omega_3-\alpha_1\gamma_4\omega_3-\alpha_3\gamma_1\omega_4+\alpha_1\gamma_3\omega_4+\alpha_3\gamma_3\omega_1\right]L^3, \\
-1+\left[\vartheta_1-\gamma_3+\omega_4\right]L+\left[\vartheta_1\gamma_3+\vartheta_4\omega_1-\vartheta_3\omega_1-\gamma_4\omega_3-\vartheta_1\omega_4+\gamma_3\omega_4\right]L^2 \\
+\left[\vartheta_4\gamma_3\omega_1-\vartheta_3\gamma_4\omega_1-\vartheta_4\gamma_1\omega_3+\vartheta_1\gamma_4\omega_3+\vartheta_3\gamma_1\omega_4-\vartheta_1\gamma_3\omega_4\right]L^3, \\
\alpha_3L+\left[\vartheta_3\alpha_2-\vartheta_1\alpha_3+\alpha_4\omega_3-\alpha_3\omega_4\right]L^2+\left[\vartheta_3\alpha_4\omega_1+\vartheta_4\alpha_1\omega_3-\vartheta_1\alpha_4\omega_3\right. \\
\left.-\vartheta_3\alpha_1\omega_4+\vartheta_1\alpha_3\omega_4-\vartheta_4\alpha_3\omega_1\right]L^3, \\
-\alpha_4L+\left[\vartheta_1\alpha_4-\vartheta_4\alpha_1-\alpha_4\gamma_3+\alpha_3\gamma_4\right]L^2+\left[\vartheta_4\alpha_3\gamma_1-\vartheta_3\alpha_4\omega_1-\vartheta_4\alpha_1\gamma_3\right. \\
\left.+\vartheta_1\alpha_4\gamma_3+\vartheta_3\alpha_1\omega_4-\vartheta_1\alpha_3\gamma_4\right]L^3.
\end{array}
$$

Row 3:

$$
\begin{array}{l}
\gamma_1L+\left[\alpha_1\gamma_2-\alpha_2\omega_1+\gamma_4\omega_1-\gamma_1\omega_4\right]L^2 \\
+\left[\alpha_4\gamma_2\omega_1-\alpha_2\gamma_4\omega_1-\alpha_4\gamma_1\omega_2+\alpha_1\gamma_4\omega_2+\alpha_2\gamma_1\omega_4-\alpha_1\gamma_2\omega_4\right]L^3, \\
\gamma_2L+\left[\vartheta_2\gamma_1-\vartheta_1\gamma_2+\gamma_4\omega_2-\gamma_2\omega_4\right]L^2+\left[\vartheta_4\gamma_1\omega_2-\vartheta_4\gamma_2\omega_1+\vartheta_2\gamma_4\omega_1\right. \\
\left.-\vartheta_1\gamma_4\omega_2-\vartheta_2\gamma_1\omega_4+\vartheta_1\gamma_2\omega_4\right]L^3, \\
1-\left[\vartheta_1+\alpha_2+\omega_4\right]L+\left[\vartheta_1\alpha_2+\vartheta_4\omega_1-\vartheta_2\alpha_1-\alpha_4\omega_2+\vartheta_1\omega_4+\alpha_2\omega_4\right]L^2 \\
+\left[\vartheta_4\alpha_2\omega_1-\vartheta_2\alpha_4\omega_1-\vartheta_4\alpha_1\omega_2+\vartheta_1\alpha_4\omega_2+\vartheta_2\alpha_1\omega_4-\vartheta_1\alpha_2\omega_4\right]L^3, \\
\vartheta_4L+\left[\vartheta_4\gamma_1+\alpha_4\gamma_2-\vartheta_1\gamma_4-\alpha_2\gamma_4\right]L^2+\left[\vartheta_2\alpha_4\gamma_1-\vartheta_4\alpha_2\gamma_1+\vartheta_4\alpha_1\gamma_2\right. \\
\left.-\vartheta_1\alpha_4\gamma_2-\vartheta_2\alpha_1\gamma_4+\vartheta_1\alpha_2\gamma_4\right]L^3.
\end{array}
$$

Row 4:

$$
\begin{array}{l}
-\omega_1L+\left[\alpha_2\gamma_1-\gamma_3\omega_1-\alpha_1\omega_2+\gamma_1\omega_3\right]L^2 \\
+\left[\alpha_2\gamma_3\omega_1+\alpha_3\gamma_1\omega_2-\alpha_1\gamma_3\omega_2-\alpha_2\gamma_1\omega_3+\alpha_1L\gamma_2L\omega_3L-\alpha_3L\gamma_2L\omega_1\right]L^3, \\
-\omega_2L+\left[\vartheta_1\omega_2-\vartheta_2\omega_1+\gamma_3\omega_2+\gamma_2\omega_3\right]L^2+\left[\vartheta_3\gamma_2\omega_1-\vartheta_2\gamma_3\omega_1+\vartheta_3\gamma_1\omega_2\right. \\
\left.+\vartheta_1\gamma_3\omega_2+\vartheta_2\gamma_1\omega_2-\vartheta_1\gamma_2\omega_3\right]L^3, \\
\omega_3L+\left[\vartheta_3\omega_1+\alpha_3\omega_2-\vartheta_1\omega_3-\alpha_2\omega_3\right]L^2 \\
+\left[\vartheta_2\alpha_3\omega_1-\vartheta_3\alpha_2\omega_1+\vartheta_3\alpha_1\omega_2-\vartheta_1\alpha_3\omega_2-\vartheta_2\alpha_1\omega_3-\vartheta_1\alpha_2\omega_3\right]L^3, \\
-1+\left[\vartheta_1-\gamma_3+\alpha_2\right]L+\left[\vartheta_2\alpha_1-\vartheta_1\alpha_2-\vartheta_3\gamma_1-\alpha_3\gamma_2+\alpha_2\gamma_3+\vartheta_1\gamma_3\right]L^2 \\
+\left[\vartheta_3\alpha_2\gamma_1-\vartheta_2\alpha_3\gamma_1-\vartheta_3\alpha_1\gamma_2+\vartheta_1\alpha_3\gamma_2+\vartheta_2\alpha_1\gamma_3-\vartheta_1\alpha_2\gamma_3\right]L^3.
\end{array}
$$

The highest lag (L) on the right-hand-side is $4th$ order in S. From the adjoint matrix we see that the highest lag of the $X's$ and ϵ is order 3.

9 Appendix B

Instrument set for single fourth-order equation in first-differenced form

$$\begin{pmatrix} S_{i1} & 0 & 0 & \dots & \dots & 0 & \dots & \dots & 0 & \Delta X_{i6} & \Delta X_{i5} & \Delta X_{i4} & \Delta X_{i3} \\ 0 & S_{i1} & S_{i2} & \dots & \dots & 0 & \dots & \dots & 0 & \Delta X_{i7} & \Delta X_{i6} & \Delta X_{i5} & \Delta X_{i4} \\ \vdots & \vdots & \vdots & \ddots & & \vdots & \ddots & & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & & \ddots & \vdots & & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \dots & S_{i1} & \dots & \dots & S_{iT-5} & \Delta X_{iT} & \Delta X_{iT-1} & \Delta X_{iT-2} & \Delta X_{iT-3} \end{pmatrix}$$