Semi-automatic Non-linear Model Selection

Jennifer L. Castle
Magdalen College and
Institute for New Economic Thinking at the Oxford Martin School, Oxford University, UK,

David F. Hendry*
Economics Department and
Institute for New Economic Thinking at the Oxford Martin School, Oxford University, UK

May, 2013

Abstract

We consider model selection for non-linear dynamic equations with more candidate variables than observations, based on a general class of non-linear-in-the-variables functions, addressing possible location shifts by impulse-indicator saturation. After an automatic search delivers a simplified congruent terminal model, an encompassing test can be implemented against an investigator's preferred non-linear function. When that is non-linear in the parameters, such as a threshold model, the overall approach can only be semi-automatic. The method is applied to re-analyze an empirical model of real wages in the UK over 1860–2004, updated and extended to 2005–2011 for forecast evaluation.

JEL classifications: C51, C22.
KEYWORDS: Non-linear Models; Location Shifts; Model Selection; *Autometrics*; Impulse-indicator Saturation; Step-indicator Saturation.

Contents

*Financial support from the Open Society Foundations and the Oxford Martin School is gratefully acknowledged, as are helpful comments from Michael P. Clements, Jurgen A. Doornik, Neil R. Ericsson, Niels Haldrup, Grayham E. Mizon and two anonymous referees.

1 Introduction

The problems confronting the selection of empirical non-linear models are legion. First and foremost is formulating the correct member from the infinite class of potential non-linear functions that could describe the economic reality. For aggregate data, one can at best hope for good approximations that capture the main non-linearities in a relatively constant way. Next, many non-linear in the variables functions are also non-linear in the parameters, necessitating iterative estimation algorithms which are probably too slow to implement within a model selection framework. Most aggregate economic time series are also non-stationary in levels, both from stochastic trends and structural breaks of various kinds. The latter can often be approximated by non-linearities, and conversely, exacerbating the difficulties of selection. Worse, an incorrect choice can be damaging for forecasting, wrongly extrapolating a non-existent shift, or a spurious non-linearity, into a future period. Moreover, all the usual specification and selection issues remain, including the appropriate set of relevant variables, their correct functional forms and lag lengths, and handling location shifts and outliers with possible concerns about the endogeneity of contemporaneous variables and measurement accuracy. The last two can be handled in principle using the instrumental variables equivalents of the methods we discuss, so we will not otherwise address those issues here other than checking the exogeneity of contemporaneous conditioning variables.

Model selection commencing from a general class of non-linear-in-the-variables functions which is then simplified to a congruent terminal model, must be semi-automatic for four reasons. First, there are almost certainly going to be more candidate variables (N) in total than observations (T), necessitating an initial automatic simplification. Secondly, the non-linearities found during this search process will usually only be an approximation to the 'best parsimonious' non-linear representation for any realistic data generating process (DGP). Thirdly, a dynamically unstable relation might be selected, which needs to be checked by an investigator after selection. Fourthly, a post-search encompassing test is required of the terminal model resulting from the search against an investigator's preferred function when that specification is non-linear in the parameters.

Correlations between relevant variables require that they all be included jointly, a seemingly impossible task when $N > T$. However, resorting to including only a small subset is bound to lead to model mis-specification and inconsistent parameter estimates, as well as potential non-constancies (see Hendry, 2009). This Gordian knot has got to be cut in one swoop, rather than slowly unravelled. Like Alexander's supposed solution, a human is up to this task only when armed with the appropriate tool, which here is a computer with automatic model selection software that can handle very large numbers of potential explanatory variables. We will use *Autometrics* (see Doornik, 2009a, and Castle, Doornik and Hendry, 2011), though other automatic approaches that can handle more variables than observation are doubtless applicable, such as RETINA: see Perez-Amaral, Gallo and White (2003, 2005), and Castle (2005).

The structure of the chapter is as follows. Section 2 considers using non-linear models of regime shifts. §2.1 examines how well systematic shifts are captured by a first-order threshold autoregressive model (denoted TAR(1)), extended in §2.2 to a logistic smooth transition autoregression (LSTAR), with the findings summarized in §2.3, then §2.4 considers forecasting from an LSTAR. Section 2 bears directly on the empirical application in section 4, where non-linear specifications that model non-linearities, breaks, outliers, and regime shifts are evaluated. Section 3 briefly discusses model selection when there are more variables than observations. §3.1 discusses testing for non-linearity then §3.2 describes some non-linear approximations based on polynomials of principal components; §3.3 addresses how multiple breaks may be detected using impulse-indicator saturation (IIS) as a part of model selection; and §3.4 discusses approximating a smooth transition autoregression. The resulting general formulation for facilitating model selection is presented in §3.5. Section 4 provides an empirical application to real wages in the UK over the past century and a half, re-analyzing Castle and Hendry (2009), updated and extended to 2005–2011 for forecast evaluation. §4.1 describes the data and theory; §4.2 the re-estimation of the previous non-linear model; §4.3 the approximating non-linear model, leading to a locally nesting non-linear model in §4.4; §4.5 estimates an LSTAR model; and §4.6 considers an alternative non-linear model suggested by Nielsen (2009) using interactive regime-shift dummies. Encompassing tests are computed in §4.7, but no model is found to encompass all the others, so all the forms of non-linearity considered approximate the non-linear reaction of real wages to inflation, confirming it is an important empirical phenomenon. Section 5 then reselects using step-indicator saturation (SIS: see Doornik, Hendry and Pretis, 2013) on a general equation which embeds the two equations in §4.4 and §4.6. Section 6 tests the super exogeneity of the conditioning variables in §4.4 using IIS, and in the model of §5 using SIS. Section 7 presents forecasts for both the growth rate and the level of real wages for the models in §4.4, §4.6 and §5 on the extended data over the problematic 'Great Recession' sample 2005–2011. Section 8 concludes. The Appendix records detailed data definitions.

2 Non-linear models for structural shifts

In this section we investigate the ability of non-linear models, in the form of threshold and transition specifications, to characterize regime shifts—changes with sufficient regularities that regimes are revisited—as against structural breaks, which are changes in the parameters of the system (see e.g., Hendry and Mizon, 1998). Our approach aims to detect both, by modelling regime shifts at the same times as allowing for breaks. Non-linearities in the form of regime shifts in the DGP would appear as structural breaks in linear-in-variables approximations. This motivates the application of IIS (discussed in §3.3) to linear models, where breaks matter substantively, and when selecting non-linear models, where indicators should not be needed if apparent shifts are indeed captured by the non-linearity, while at the same time protecting against a spurious non-linear fit approximating genuine breaks.

We begin by analysing the probabilities of switching regimes jointly with the magnitudes of the regime shifts in a threshold autoregressive model of order one (TAR(1)), to investigate detecting shifts in a simple model of regime change. We then consider the more realistic functional form of an LSTAR model in a small Monte Carlo. Estimation difficulties result from the inherent trade-off between the frequencies of regime shifts and the magnitudes of the shifts between regimes. Estimation requires enough obervations in all regimes, but the regimes need to be sufficiently distinct. We then look at the forecast performance of the LSTAR model compared to a linear first-order autoregressive process, AR(1). We confirm that it is often difficult to beat forecasts from the AR(1) model on a root mean square forecast error (RMSFE) criterion (see e.g., Clements and Krolzig, 1998). Unfavourable cases for LSTAR include situations when the mean shift between regimes is small, so a linear approximation is reasonable, or when the frequency of regime shifts is low, so a linear approximation performs well in small samples. Nevertheless, the empirical application in section 4 finds the non-linear model forecasts are superior. One possible explanation is that a non-linear in the variables model that uses interaction dummies to capture the regime shifts is more flexible and easier to estimate than the non-linear in the parameters LSTAR specification. The empirical exercise in §4 also finds that the linear in the parameters approximation to the LSTAR specification described in §3.4 is a feasible alternative, as it is not encompassed by the LSTAR.

2.1 Shifts captured by a threshold autoregressive model (TAR)

We first analyze estimation issues in regime-shift models by considering a TAR model of the form:

$$x_t = \sum_{1\leq i\leq m} \left(\beta_{i,0} + \beta_{i,1}x_{t-1} + \ldots + \beta_{i,p}x_{t-p} + \sigma_i\eta_t\right) \mathsf{I}\left(c_{i-1} \leq x_{t-d} \leq c_i\right) \tag{1}$$

where c_i are the thresholds, p is the longest lag, m is the number of regimes and d is the delay: see Tong (1983). We consider a delay of 1 period, $d = 1$, $m = 2$ regimes and $p = 1$, generating two regimes (upper and lower), in each of which we analyze the process as an autoregressive process of order 1, then simulate the TAR(1). Such an analysis ignores the dynamics from the previous regime shift, focusing on the properties of a stationary Gaussian AR(1) process within each regime, to ascertain the difficulties of observing enough data in each regime to sustain accurate estimation.

Let an AR(1) process in $\{y_t\}$ commence in a 'lower' regime, defined by $y_{t-1} \leq c$:

$$y_{t|\{y_{t-1}\leq c\}} = \mu + \rho y_{t-1} + \epsilon_t \tag{2}$$

where $\epsilon_t \sim \mathsf{IN}[0, 1]$. We use parameter values of $\mu = 0$ and $\rho = 0.8$, giving a realistic degree of persistence for macroeconomic time series, which results in $\mathsf{V}[y_t] = \sigma_y^2 = \frac{1}{(1-\rho^2)} = 2.78$ within that regime. The 'upper' regime with $\mu^* > \mu$ is generated by:

$$y_{t|\{y_{t-1}>c\}} = \mu^* + \rho y_{t-1} + \epsilon_t \tag{3}$$

where the error has the same distribution in both regimes. To calculate a shift in the mean of the process of magnitude $\lambda\sigma_y$, where $\lambda = 1, ..., 5$, we require $\mathsf{E}[y_t]$ to shift from 0 in (2) to $\frac{\mu^*}{(1-\rho)}$ in (3). Hence, we let $\mu^* = \frac{1}{3}, \frac{2}{3}, 1, \frac{4}{3}$ and $\frac{5}{3}$, to create shifts in mean of 1 to 5 standard deviations between regimes.

A 5% probability of a shift in the right-hand tail of the distribution of $y_{t|\{y_{t-1}\leq c\}}$ can be calculated as $\mathsf{P}\left(\frac{y_{t|\{y_{t-1}\leq c\}}-\mu}{\sigma_y} > 1.645\right) \approx 0.05$, since $\frac{y_{t|\{y_{t-1}\leq c\}}-\mu}{\sigma_y} \sim \mathsf{N}[0, 1]$ within the regime, and hence the threshold $c = (1.645 \times \sigma_y) = 2.74$ will deliver a 5% probability of shifting to the upper regime. Table 1 records a range of regime-shift probabilities for varying thresholds, given the parameters specified which determine $\sigma_y^2 = 2.78$.

The table demonstrates that there is a trade-off between the magnitude of a regime shift and the probability of a shift. A large magnitude implies a small probability of shifting again, once in the new regime, such that the number of observations in one of the regimes will likely be small and estimation difficult. A smaller mean shift implies that there is more chance of switching between regimes, which should reduce the parameter estimation uncertainty, but a smaller regime shift will be more difficult to detect, so a linear representation may prove preferable. To investigate this, we calculate the probability of switching back to the initial (lower) regime once in the upper regime. Commencing with (2), a threshold

Threshold	Probability of regime shift
3.877	1%
2.741	5%
2.135	10%
1.402	20%

Table 1: Thresholds for the probability of a shift in the right-hand tail of the initial lower regime to the upper regime.

of $c = 2.74$ will give a 5% probability of a break in the right-hand tail. Consider a regime shift of $2\sigma_y$ so the intercept shifts from $\mu = 0$ to $\mu^* = 2/3$, resulting in the unconditional mean $\mathsf{E}[y_t]$=0 shifting to $\mathsf{E}\left[y_{t|\{y_{t-1}>c\}}\right] = \frac{\mu^*}{1-\rho} = 10/3$. Once in the upper regime (3), the probability of returning to the lower regime can be calculated by considering the left-hand tail:

$$\mathsf{P}\left(y_{t|\{y_{t-1}>c\}} \leq 2.74\right) \tag{4}$$

This is computed by rescaling to the standard normal distribution:

$$\mathsf{P}\left(y_t \leq \frac{c - \mathsf{E}\left[y_{t|\{y_{t-1}>c\}}\right]}{\sigma_y}\right) = \mathsf{P}\left(y_t \leq \frac{2.74 - 10/3}{5/3}\right) \simeq 0.361 \tag{5}$$

so the probability of switching back to the lower regime is approximately 36%.

		Magnitude of mean shift to new regime				
p	c	$1\sigma_y$	$2\sigma_y$	$3\sigma_y$	$4\sigma_y$	$5\sigma_y$
1%	3.88	91%	63%	25%	4.7%	0.4%
5%	2.74	74%	36%	8.8%	0.9%	0.1%
10%	2.14	61%	24%	4.3%	0.3%	0.0%
20%	1.40	44%	12%	1.5%	0.0%	0.0%

Table 2: The probability p of a shift from the upper regime back to the lower regime, where c is the corresponding threshold value when $\mu = 0$, $\rho = 0.8$, $\sigma_\epsilon = 1$.

Table 2 records these probabilities for a range of mean shift magnitudes and thresholds. The results are dependent on the magnitude of the regime shift and the threshold value (which corresponds to the probability, p, of a regime shift from the lower to upper regime). When the mean shift is large, the probability of crossing the threshold again to return to the initial regime is low. Likewise, when there is a high probability of switching, the threshold will be small. There is a trade-off between having sufficiently distinct regimes that are of a substantive magnitude to estimate the model, whilst ensuring the mean shifts are not too large so the process 'gets stuck' in one regime. This is a small-sample problem as, with enough data, estimation of the two regimes model should be feasible, assuming that the DGP is known.

2.2 Logistic smooth transition autoregression (LSTAR)

Rather than a jump at the threshold c as in §2.1, consider an LSTAR formulation:

$$y_t = \mu + \rho y_{t-1} + \mu^* \left[1 + \exp\left(-\gamma\left(\frac{y_{t-1} - c}{\sigma_y}\right)\right)\right]^{-1} + \epsilon_t \tag{6}$$

developed by Maddala (1977), Granger and Teräsvirta (1993), and Teräsvirta (1994).[1] In (6), γ determines the rapidity of the transition from 0 to 1 as a function of the transition variable, y_{t-1} with standard deviation σ_y, and c determines the transition point. Both γ and c must be estimated, as in Teräsvirta (1994) and Franses and Van Dijk (2000).[2] Estimation of γ is difficult, as the likelihood function is not well behaved even with a known functional form and $\gamma > 0$ as an identifying restriction: see Granger and Teräsvirta (1993), p.123. Let:

$$F(z_t) = (1 + \exp\{-z_t\})^{-1} \tag{7}$$

where:

$$z_t = \gamma\left(\frac{y_{t-1} - c}{\sigma_y}\right) \tag{8}$$

As $F(\cdot)$ is the logistic cdf, an upper bound on z_t of approximately 10 can be deduced from Chebyshev's inequality, $\Pr(z_t \geq 10) \leq 0.00005$, suggesting an upper bound on $\widehat{\gamma}$ of around 5. For $\widehat{\gamma} \geq 5$, the transition function approximates a two regime-switching process, so (6) simplifies to a switching autoregression. If $\widehat{\gamma}$ is close to zero, the increased uncertainty regarding the regime increases the uncertainty of other parameter estimates, but this is less likely after ensuring that the relationship is non-linear.

To illustrate, we set $\gamma = 3$ and generate $T = 100$ observations, after discarding an initial 100 observations. Thus, the beginning of the sample could lie in either the upper or lower regime. Table 3 records the correlation between the LSTAR and the TAR model for varying γ for $M = 10,000$ replications. We report the correlation coefficient for 3 different shift magnitudes ($1\sigma_y$, $3\sigma_y$ and $5\sigma_y$) and for various shift probabilities (1% to 20%). Increasing γ increases the correlation between the LSTAR and TAR as the speed of transition is increased, and by $\gamma = 5$, the smooth transition is almost equivalent to a step shift. There is a non-linear relationship between the size of shift, probability of shift, and the correlation between the LSTAR and TAR models. For small shifts (i.e., σ_y), increasing the probability of a shift reduces their correlation, but as the magnitude of the shift increases, the correlation first falls and then increases. The lowest correlation between the two models occurs when the shift is large but the probability of switching is low, or when the shift is moderate but the probability of a shift is moderate too. In these cases, the occurrence of shifts is likely to be higher, and the divergence between the two models increases as the smooth transition component has a larger impact.

We next investigate the probability of detecting a shift with a Monte Carlo experiment, where a shift in the LSTAR model is any realisation that exceeds the threshold, c. The transition function for one draw at $\gamma = 4$ is recorded in Figure 1 (the small volatility in the LSTAR function close to 0 or 1 does not count as a transition). Observe the divergent behaviour of the two transition functions at the beginning of the sample (even though the initial 100 observations are discarded). It is possible to get very different behaviour from the two transitions depending on past values, but the correlations indicate that this is rare.

We simulate 10,000 replications of the DGP (6) for a sample size of 100, using a value of $\gamma = 3$ for all replications. Table 4 records the number of observations in the upper regime, the number of regime shifts on average, the number of shifts from the lower to the upper regime, and the average number of observations in the upper regime before a switch. The threshold parameter takes four values, corresponding to a regime shift probability from the lower to the upper regime of 1%, 5%, 10% and 20%. Three mean shift sizes are also examined: $1\sigma_y$, $3\sigma_y$ and $5\sigma_y$. The LSTAR model estimates more regime shifts on average than the TAR model. For small shifts, the number of regime switches increases as the probability of a regime shift increases, but for moderate shifts this is not monotonic. As the probability of a mean shift increases, the threshold falls and hence the probability of switching back is lower for larger

[1] Variations result in other regime-switching models including smooth-transition autoregressions (STAR), see Chan and Tong (1986) and Luukkonen, Saikkonen and Teräsvirta (1988); TAR as above; switching regression models, see Quandt (1983); and exponential autoregression models (EAR), see Priestley (1981).

[2] A set of non-linear functions could be generated for a range of values of γ, c and included in the initial general model, with an automatic search procedure like *Autometrics* used to select the functions with the most appropriate values.

c		$\gamma = 1$	$\gamma = 3$	$\gamma = 5$	$\gamma = 10$	$\gamma = 100$
3.87	$1\sigma_y$	0.9983	0.9996	0.9998	0.9999	1.0000
	$3\sigma_y$	0.9527	0.9796	0.9863	0.9912	0.9987
	$5\sigma_y$	0.7026	0.8475	0.9099	0.9502	0.9932
2.74	$1\sigma_y$	0.9956	0.9986	0.9992	0.9996	0.9999
	$3\sigma_y$	0.8606	0.9186	0.9415	0.9630	0.9933
	$5\sigma_y$	0.8867	0.9333	0.9530	0.9705	0.9960
2.14	$1\sigma_y$	0.9922	0.9975	0.9987	0.9993	0.9999
	$3\sigma_y$	0.8685	0.9233	0.9454	0.9661	0.9936
	$5\sigma_y$	0.9666	0.9842	0.9878	0.9926	0.9989
1.40	$1\sigma_y$	0.9873	0.9963	0.9980	0.9990	0.9998
	$3\sigma_y$	0.9349	0.9677	0.9779	0.9858	0.9969
	$5\sigma_y$	0.9935	0.9991	0.9994	0.9996	0.9999

Table 3: Correlation between TAR(1) and LSTAR(1) for $T = 100$

mean shifts. When the mean shifts are large, the process tends to stay in one regime. Even for moderate breaks, there are so few regime shifts that estimation could prove difficult.

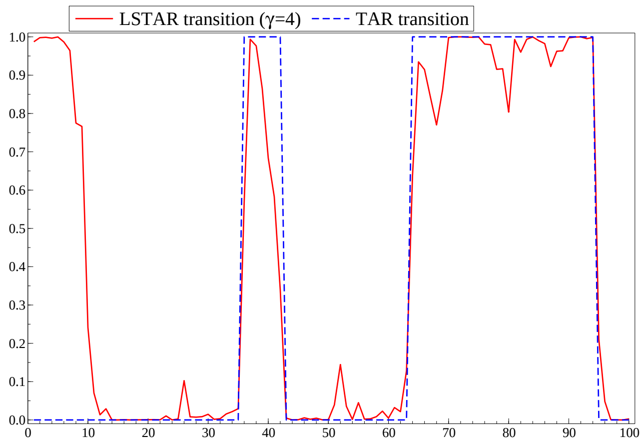

Figure 1: Transition functions for TAR and LSTAR models: $\gamma = 4$

Finally, we investigate the impact of the occurrence of regime switches on estimation of the LSTAR model. Table 5 reports the equation standard error and the Schwarz information criterion *SIC* (see Schwarz, 1978) for the correctly specified LSTAR model and for a mis-specified AR(1) process (which would be correctly specified if there were no regime shifts in the in-sample period). When the process is in the upper regime, the linear intercept is given by $\mu + \mu^*$. As the likelihood function is often flat, convergence to extreme values can occur. Hence, we exclude any draw that either does not converge or that results in any of the intercept or autoregressive parameters (i.e. μ, μ^* or ρ) exceeding 10 in absolute value. We record the number of excluded replications as errors. 1,000 replications are undertaken.

The estimates for the LSTAR model are poor, reflected in the large mean equation standard errors, and the huge Monte Carlo standard deviations on both the equation standard error and SIC, which highlight that some draws lead to very poor estimates. The equation standard errors of the mis-specified AR(1) model are close to the DGP standard error of unity regardless of the shift probability or magnitude, suggesting that few shifts are generated by this DGP. Thus, estimation issues may hinder the use of the LSTAR model in small samples when shifts are not large and frequent. The estimates seem overly

c		3.87		2.74		2.14		1.40	
		TAR	LSTAR	TAR	LSTAR	TAR	LSTAR	TAR	LSTAR
$\mathbf{1\sigma_y}$	No. obs upper	1.38	1.51	7.68	8.20	16.16	16.84	33.15	33.66
	No. shifts	1.25	1.50	5.05	5.96	8.36	9.81	12.32	14.26
	No. shifts upper	0.63	0.75	2.53	2.98	4.18	4.91	6.16	7.13
	Ave. length upper	2.13	1.96	3.01	2.73	3.84	3.41	5.41	4.73
$\mathbf{3\sigma_y}$	No. obs upper	5.32	5.81	37.81	38.28	67.90	67.65	91.05	90.36
	No. shifts	1.26	2.23	3.55	6.26	3.40	5.89	1.76	3.13
	No. shifts upper	0.63	1.12	1.78	3.14	1.71	2.95	0.88	1.57
	Ave. length upper	7.66	4.99	18.12	11.97	27.36	20.42	37.30	31.40
$\mathbf{5\sigma_y}$	No. obs upper	42.76	55.96	92.67	97.83	99.06	99.81	99.97	99.99
	No. shifts	0.49	1.24	0.21	0.23	0.05	0.05	0.01	0.01
	No. shifts upper	0.33	0.70	0.16	0.13	0.04	0.02	0.00	0.00
	Ave. length upper	26.70	23.28	33.57	38.02	34.45	43.68	41.86	36.10

Table 4: Probability of a shift in the TAR and LSTAR models. ($\gamma = 3$)

c		3.87		2.74		2.14		1.40	
		LSTAR	AR	LSTAR	AR	LSTAR	AR	LSTAR	AR
$\mathbf{1\sigma_y}$	$\widehat{\sigma}$	11.164 (105.18)	1.001 (0.07)	13.402 (171.76)	1.004 (0.07)	10.791 (82.39)	1.003 (0.07)	24.602 (260.56)	1.002 (0.07)
	SIC	0.942 (2.06)	0.070 (0.14)	1.085 (2.15)	0.075 (0.14)	1.291 (2.32)	0.072 (0.14)	1.596 (2.64)	0.070 (0.14)
	No. errors	23.6%		20.2%		17.3%		18.4%	
$\mathbf{3\sigma_y}$	$\widehat{\sigma}$	19.798 (229.32)	1.005 (0.07)	10.266 (56.77)	1.011 (0.07)	40.355 (772.92)	1.011 (0.07)	7.847 (57.62)	1.003 (0.07)
	SIC	1.066 (2.27)	0.076 (0.14)	1.663 (2.50)	0.089 (0.14)	1.774 (2.71)	0.087 (0.15)	1.086 (2.15)	0.073 (0.15)
	No. errors	18.7%		20.8%		19.2%		13.3%	
$\mathbf{5\sigma_y}$	$\widehat{\sigma}$	5.809 (38.55)	1.008 (0.08)	2.555 (11.29)	1.000 (0.07)	7.432 (112.61)	0.999 (0.07)	2.433 (23.10)	0.999 (0.07)
	SIC	1.020 (1.94)	0.083 (0.15)	0.532 (1.33)	0.067 (0.14)	0.495 (1.44)	0.066 (0.14)	0.399 (1.08)	0.064 (0.14)
	No. errors	14.0%		7.4%		6.5%		5.6%	

Table 5: Equation standard error and SIC for the LSTAR(1) and AR(1) models, with Monte Carlo standard deviations reported in parentheses. ($\gamma = 3$)

dependent on the starting values for the optimisation, which here were the actual DGP values. Table 6 compares these results to initial values of 0 and 1 for all parameters for a 5% probability of a shift and a shift magnitude of $3\sigma_y$. The mean equation standard error is substantially increased by these initial conditions, again highlighting difficulties with estimating the LSTAR model.

2.3 In-sample summary

The numbers and magnitudes of shifts are fundamental to the estimation of threshold models. In the event that shifts are rare, threshold values will be large, implying the probability of switching regime will be low. On the other hand, if the probability of a shift is high, the threshold will be low and if the shift magnitude is large, the probability of switching back to the initial regime will be low. Estimation of the LSTAR model seems difficult because the likelihood function is not always well behaved. The Monte Carlo evidence suggests estimating the DGP is substantially harder than approximating it by an AR(1) process, regardless of the shift probability or size. These results may be due to small sample sizes which imply a lack of shifts.[3]

	Initial Conditions		
	DGP	0	1
$\widehat{\sigma}$	10.3	1397	122.1
	(56.77)	(15927)	(1686)
SIC	1.663	7.212	2.476
	(2.50)	(4.61)	(3.26)
No. errors	20.8%	4.6%	31.2%

Table 6: The impact of initial values on the estimates of the LSTAR model (for a shift probability of 5% with a magnitude of $3\sigma_y$).

2.4 Forecasting using the LSTAR model

In this section, building on Castle, Fawcett and Hendry (2011), we evaluate the forecast performance of the LSTAR model for a simple DGP to provide guidance on interpreting the subsequent empirical results: general discussions of forecasting with LSTAR and other non-linear models are provided in Lundbergh and Teräsvirta (2002) and Kock and Teräsvirta (2011). The forecasting exercise considers two sample sizes; $T = 100$ and $T = 1000$, where $H = 20$ 1-step ahead forecasts are computed for the sample size of 100 and $H = 200$ forecasts are computed when $T = 1000$. The DGP is given by equation (6), with $\gamma = 3$. 1000 replications were undertaken and forecasts were computed using in-sample parameter estimates from the initial conditions set at the DGP values. Draws in which the parameter estimates were extreme were discarded, but a number of draws were still erratic, leading to large RMSFEs. Hence, we report the percentage of draws in which the RMSFE of the LSTAR model was less than that of a benchmark AR(1) forecast. If the transition function is 0 or 1 over the entire in-sample period, the LSTAR model simplifies to an AR(1) process so when regime shifts are infrequent, many draws produce identical forecasts from the two models. Thus, Table 7 reports the proportion of draws in which the RMSFEs for the LSTAR model were equal to the AR(1) model, or lower than those of the AR(1) model. We also compared peformance to a random walk, but both LSTAR and AR(1) were superior.

For small regime shifts (σ_y), it is difficult to beat the AR(1) model—less than 40% of draws deliver better forecasts. Increasing the sample size does not yield greatly improved forecast performance either, so the estimated correctly specified model remains a poor representation of the DGP. The probability

[3]Nevertheless, we focus on the LSTAR, rather than the TAR, model in the subsequent analysis as the more general model.

c	3.87		2.74		2.14		1.40	
T	100	1000	100	1000	100	1000	100	1000
$\mathbf{1\sigma_y}$								
=AR(1)	0.08	0.26	0.04	0.18	0.04	0.16	0.03	0.16
<AR(1)	0.35	0.27	0.38	0.36	0.37	0.40	0.36	0.38
$\mathbf{3\sigma_y}$								
=AR(1)	0.07	0.07	0.03	0.03	0.03	0.03	0.17	0.05
<AR(1)	0.40	0.70	0.47	0.81	0.50	0.83	0.39	0.80
$\mathbf{5\sigma_y}$								
=AR(1)	0.23	0.02	0.62	0.45	0.71	0.69	0.81	0.82
<AR(1)	0.41	0.93	0.17	0.34	0.13	0.13	0.08	0.07

Table 7: Percentages of draws in which the LSTAR model RMSFE is equal to or less than that of the AR(1) model.

that the non-linear model is identical to the AR(1) model increases with sample size. Hence, with small regime changes, even large sample sizes do not pick up the non-linearity.

With moderate sized regime shifts ($3\sigma_y$), it is easier to distinguish between the LSTAR and AR(1) model regardless of the probability of a switch. At small samples, the LSTAR forecast performance is poor relative to the AR(1) model, but at larger sample sizes, the correct model performs much better.

Finally, for large shifts, the LSTAR model often coincides with the AR(1) process for the given sample, particularly as the probability of a regime shift increases: once in the upper regime, the process is likely to remain there, so an AR(1) model is then correct. Although the LSTAR nests the AR(1), so remains correctly specified, it is over-parameterized, and there is a lack of identification.

3 Model selection with more variables than observations

The difficulties just described are compounded when a multi-path search procedure like *Autometrics* is used: iterative estimation of such non-linear-in-parameters models during multi-path search over other variables, lags and possible breaks seems infeasible. Thus, after describing a test for non-linearity in §3.1, we consider approximating non-linearites by polynomials (§3.2), and impulse-indicator saturation (IIS) for tackling multiple location shifts (§3.3), then develop an approximation to an LSTAR model (§3.4).

3.1 Testing for non-linearity

An index test for non-linearity can be computed to determine whether the initial linear specification should include non-linear functions. Castle and Hendry (2010) provide details of the test, in which principal components of the set of possible linear regressors are computed and their non-linear functions are jointly tested. Let $\mathbf{x}_t$ denote the set of candidate regressors where $\mathbf{x}_t \sim \mathsf{D}_n[\boldsymbol{\mu}, \boldsymbol{\Omega}]$ and $\boldsymbol{\Omega}$ is their symmetric, positive-definite variance-covariance matrix. Factorize $\boldsymbol{\Omega} = \mathbf{H}\boldsymbol{\Lambda}\mathbf{H}'$, where $\mathbf{H}$ is the matrix of eigenvectors of $\boldsymbol{\Omega}$ and $\boldsymbol{\Lambda}$ the corresponding eigenvalues, such that $\mathbf{H}'\mathbf{H} = \mathbf{I}_n$. Since $\boldsymbol{\Lambda}^{1/2}\mathbf{H}'\boldsymbol{\Omega}\mathbf{H}\boldsymbol{\Lambda}^{1/2} = \mathbf{I}_n$, let $\mathbf{z}_t = \boldsymbol{\Lambda}^{1/2}\mathbf{H}'(\mathbf{x}_t - \boldsymbol{\mu}) \sim \mathsf{D}_n[\mathbf{0}, \mathbf{I}]$. Specify $u_{1,i,t} = z_{i,t}^2$; $u_{2,i,t} = z_{i,t}^3$; and $u_{3,i,t} = z_{i,t}e^{-|z_{i,t}|}$, such that under the null, a test of $\boldsymbol{\delta}_1 = \boldsymbol{\delta}_2 = \boldsymbol{\delta}_3 = \mathbf{0}$ in:

$$y_t = \beta_0 + \boldsymbol{\beta}'\mathbf{x}_t + \boldsymbol{\delta}_1'\mathbf{u}_{1,t} + \boldsymbol{\delta}_2'\mathbf{u}_{2,t} + \boldsymbol{\delta}_3'\mathbf{u}_{3,t} + \epsilon_t \quad (9)$$

is an exact F-test with $3n$ degrees of freedom for fixed regressors (approximately F otherwise). This compares to $\left\{\frac{n(n+1)}{2}\left(1 + \frac{(2n+1)}{3}\right) + n^2\right\}$ degrees of freedom for a general test of all squares, cubics and exponentials of the original regressors, which would often lead to more variables than observations.

3.2 Non-linear approximations

The class of non-linear-in-variables functions that might be entertained is vast. Viewed as approximations to an unknown non-linear relation, the key consideration is how closely a given specification might represent the unknown member from a potentially wide class of functions. Viewed from a model selection and estimation perspective, the important issue becomes the parsimony of that approximation, so relatively precise estimates can be obtained. There are many possible choices for the first stage, including polynomial expansions, trigonometric and hypergeometric series, and squashing functions. Third-order polynomials augmented by exponentials of the principal components (PCs) of the levels of the original variables, as used in the above index test for non-linearity, provide a low-dimensional solution when interactions between non-linear functions matter. However, absent such interactions, then non-linear functions of the individual variables may prove more parsimonious. In the empirical example below, the cross correlations between variables are relatively low (other than productivity and real wages), so we use the latter. Either way, after selecting a parsimonious terminal model from general polynomials, seeking a further reduction by an encompassing test against a theory-based form provides the second, non-automatic, stage of selection. §3.4 considers a polynomial approximation to LSTAR.

3.3 Impulse-indicator saturation

Impulse-indicator saturation includes in the set of candidate variables an impulse indicator for every observation, $1_{t=j} \forall t = 1, \ldots, T$, ensuring $N > T$. IIS is analyzed by Hendry, Johansen and Santos (2008) and Johansen and Nielsen (2009), who show that the costs of including T indicators under the null that none is relevant are low when α is set at $1/T$, namely 1% when $T = 100$ despite selecting over 100 variables. Castle, Doornik and Hendry (2012) show that IIS has good power to detect outliers, location shifts and alleviate problems of inference from fat-tailed distributions. Conversely, not handling outliers and shifts can distort inference and, especially in the context of non-linear model selection, can lead to mistaken choices of functional form (see e.g., Castle and Hendry, 2011). Section 6 considers the application of IIS to test the exogeneity of contemporaneous conditioning variables.

3.4 Approximating a smooth transition autoregression

The logistic transition function $F(z_t)$ can be approximated by a 3rd-order Taylor expansion as:

$$F(z_t) \simeq \left(\frac{1}{2} + \frac{z_t}{4} - \frac{z_t^3}{48} \right) \tag{10}$$

The z_t^2 term drops out as $\partial^2 F(z)/\partial z^2 \rfloor_{z=0} = 0$. However, a quadratic component could still be included in the model to allow for interactions like $y_{t-1}F(z_t)$. For an LSTAR like (6), this approximation delivers:

$$y_t = \mu^{**} + \rho y_{t-1} + \mu_1^* y_{t-1}^2 + \mu_2^* y_{t-1}^3 + v_t \tag{11}$$

Although the transition variable is scaled by both γ and $\widehat{\sigma}_y$, they are included in the coefficients when estimating the polynomial approximation. In a univariate setting, the mappings from the coefficients in (6) to (11) are known. Then (11), and generalizations thereof allowing for lags and location shifts, can be estimated by non-iterative methods to facilitate selection, subject to v_t being approximately distributed as $\mathsf{ID}\left[0, \sigma_v^2\right]$, which can be checked. After selection, a test of linearity in the autoregressive model (11) is a special case of the test in §3.1, namely whether any of the non-linear functions are retained:

$$\mathsf{H}_0 : \mu_1^* = \mu_2^* = 0 \tag{12}$$

Then the LSTAR model can be tested by estimating the version that entailed the approximating model, and testing the elimination of all the selected approximating terms. For the more general example considered below, let there be n relevant variables $\mathbf{x}_t$ of which $k \leq n$ were retained after selection, denoted $\mathbf{x}_t^*$, and one transition variable was selected, denoted $s_{1,t}$, then the test of the approximation to the LSTAR model would be:

$$\mathsf{H}_0 : \boldsymbol{\kappa}_2 = \boldsymbol{\kappa}_3 = \boldsymbol{\kappa}_4 = \mathbf{0} \tag{13}$$

in the encompassing regression:

$$\begin{aligned} y_t &= \boldsymbol{\kappa}_1'\mathbf{x}_t^* + \boldsymbol{\kappa}_2'\mathbf{x}_t^* s_{1,t} + \boldsymbol{\kappa}_3'\mathbf{x}_t^* s_{1,t}^2 + \boldsymbol{\kappa}_4'\mathbf{x}_t^* s_{1,t}^3 \\ &\quad + \left(\boldsymbol{\theta}'\mathbf{x}_t\right)\left(1 + \exp\left\{-\gamma\left(\frac{s_t - c}{\widehat{\sigma}_s}\right)\right\}\right)^{-1} + \eta_t \end{aligned} \tag{14}$$

This approach ensures that the model is non-linear, and checks whether the LSTAR formulation captures all of that non-linearity, which occurs only if no additional non-linearities are retained in (14).

3.5 The general formulation

Consider a local DGP of the form:

$$\psi_y(L)y_t = f\left(\psi_{z_1}(L)z_{1,t}, \ldots, \psi_{z_k}(L)z_{k,t}; \boldsymbol{\theta}\right) + \boldsymbol{\delta}'\mathbf{d}_t + \epsilon_t \;\text{ where }\; \epsilon_t \sim \mathsf{IN}\left[0, \sigma_\epsilon^2\right] \tag{15}$$

where $z_{i,t}$ denotes the set of k linear conditioning variables, $\psi_y(L)$ and $\psi_{z_i}(L), i = 1, \ldots, k$ are lag polynomials, and $\mathbf{d}_t$ are dummy variables for $t = 1, \ldots, T$, with $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Model selection must address the specification of the functional form, $f(\cdot)$, the identification of $\boldsymbol{\theta}$, the selection of the potentially relevant variables, $\mathbf{z}_t' = (z_{1,t}, \ldots z_{k,t})$ from the available candidates $(z_{1,t}, \ldots z_{K,t})$ where $K \geq k$, the lag lengths, and all outliers and shifts. We approach this problem using an extended general-to-specific methodology, whereby the initial general unrestricted model (GUM) is specified to nest (15). As the functional form is unknown and a Taylor expansion of $f(\cdot)$ around zero would result in a very rapidly increasing number of parameters as K grows, we use the approximating functions discussed above. Similarly, we use IIS to model the $\{\mathbf{d}_t\}$.

Let $w_{i,t}$ denote either the original variables, $z_{i,t}$, or their principal components, then the initial general unrestricted specification with s lags is:

$$\begin{aligned} y_t &= \sum_{i=1}^{K}\sum_{j=0}^{s}\beta_{i,j}z_{i,t-j} + \sum_{i=1}^{K}\sum_{j=0}^{s}\kappa_{i,j}w_{i,t-j}e^{-|w_{i,t-j}|} + \sum_{i=1}^{K}\sum_{j=0}^{s}\theta_{i,j}w_{i,t-j}^2 \\ &+ \sum_{i=1}^{K}\sum_{j=0}^{s}\gamma_{i,j}w_{i,t-j}^3 + \sum_{j=1}^{s}\lambda_j y_{t-j} + \sum_{i=1}^{T}\delta_i 1_{\{i=t\}} + \epsilon_t \end{aligned} \tag{16}$$

where T is the maximum available sample size. A formulation like (16) leads to $N = 4K(s+1) + s + T$ right-hand side candidate variables including lags, functional form transforms, deterministic terms (including indicator variables), so the approach is bound to generate $N > T$.

As the GUM in (16) is not feasibly estimable, it is impossible to tackle all issues jointly. However, the block sequential search discussed in Hendry and Krolzig (2005) and Doornik (2009a, 2009b) has been shown to be effective in related settings where $N > T$, so we adopt that approach below. Thus all candidate variables are included in the set to be selected over, and entered in large blocks (rather than singly as in, say, stepwise regression methods), with a record kept of which were significant at the chosen level of $\alpha\%$. Next, significant variables are combined in a further selection, where the resulting terminal model is tested against blocks of not-yet-included candidates. Hendry and Johansen (2013) extend the analysis in Hendry (2000) to show that under the null that all N variables are irrelevant, αN will be retained by chance even when $N > T$. Moreover, they show that when a theory-model is retained without selection, under the null that it is a complete and correct specification, by orthogonalizing all other variables with respect to the theory-based set, despite selecting from those, the resulting parameter estimates will be identical to those obtained from directly fitting that theory-model to data.[4] Thus, even with $N > T$, theory-based model selection can be nearly costless.

4 Empirical application

Econometric models of wage inflation have a long history, see *inter alia* Dicks-Mireaux and Dow (1959), Lipsey (1960), Phillips (1958), Sargan (1964, 1980), Godley and Nordhaus (1972), Nickell (1990), and Layard, Nickell and Jackman (1991): Henry (1982) provides a historical perspective on empirical models of wages. The specification of these models varies greatly: early models considered nominal wages, followed by models with real-wage equilibria, and finally inflation expectations were accorded a key role, becoming dominant in the 'New-Keynesian' approach to price inflation. Despite this plethora of models, there is still uncertainty as to the preferable specification of a wage inflation model, with the literature divided between the role of feed-forward versus feedback mechanisms: see the contrasting models of Castle and Hendry (2009) and the New Keynesian Phillips Curve (NKPC) models proposed by Galí and Gertler (1999) and Galí, Gertler and Lopez-Salido (2001), with a critique in Castle, Doornik, Hendry and Nymoen (2012b).

Using the example of UK real wages over the past century and a half, we demonstrate that all substantively relevant variables, dynamics, outliers and breaks, and non-linearities must be modelled jointly for a coherent empirical economic model. The same theory model that real wages are determined by the marginal product of labour underlies all the different specifications considered. However, both static and dynamic linear models without IIS provided poor statistical representations, and did not adequately capture the underlying data properties, with few variables 'significant', albeit greatly improved if augmented by IIS. Thus, outliers and shifts must be modelled for a valid statistical representation, and using IIS allows political, institutional and external events to be selected without imposing any *a priori* assumptions, using the data to determine the timing of extraneous events. Testing for non-linearity in the general linear dynamic specification pointed towards a possible non-linear relationship, so we commenced with a general non-linear approximation, then undertook selection for several non-linear functions that were linear in the parameters. We show below that an LSTAR specification which is non-linear in the parameters is a restricted version of the more general non-linear model. Non-linear functions seem important in explaining real wage growth, and those functions suggest a causal relationship with unemployment rates.

We use the semi-automated approach explained above, attempting to encompass the selected model with that reported in Castle and Hendry (2009), in which a non-linear wage-price spiral term was found to be important. The non-linear modelling also allows for regime shifts, following from Nielsen (2009). We then examine whether reductions can be made by eliminating the non-linear functions, indicators and dynamics. Encompassing tests do not allow for such reductions to be made, although super-exogeneity tests confirm the viability of analyzing single-equation equilibrium-correction mechanisms. Finally, we forecast the last 7 years of real wages on the extended data set, and find the highly parameterized non-linear model forecasts most accurately on a MSFE criterion, so parsimony need not be preferable and non-linear models can outperform linear in a forecasting context.

[4]The Hendry and Johansen (2013) method for incorporating a theory model could potentially be applied using $\widehat{y}$ from a model that is non-linear in parameters, but we do not address that here.

4.1 The data and theory

The data are annual time series for the UK over 1860–2004 based on Castle and Hendry (2009), updated and extended to 2011, providing 7 additional observations for forecasting. The data sources are detailed in the appendix. The main variables are nominal wages, w_t, and prices, p_t, in logs (shown in fig. 2a, adjusted to match means for clarity). In our analysis, these variables are assumed to be I(1), with real wages, $(w-p)_t$, also I(1) (see fig. 2b), whereas wage and price inflation, $(\Delta w_t, \Delta p_t)$, are I(0), although subject to breaks and regime changes. Wage and price inflation also cobreak, as can be seen in fig. 2c. Real-wage inflation, $\Delta(w-p)_t$, is I(0) with one large outlier in 1940 (shown in fig. 2d). This last is the dependent variable in our models.

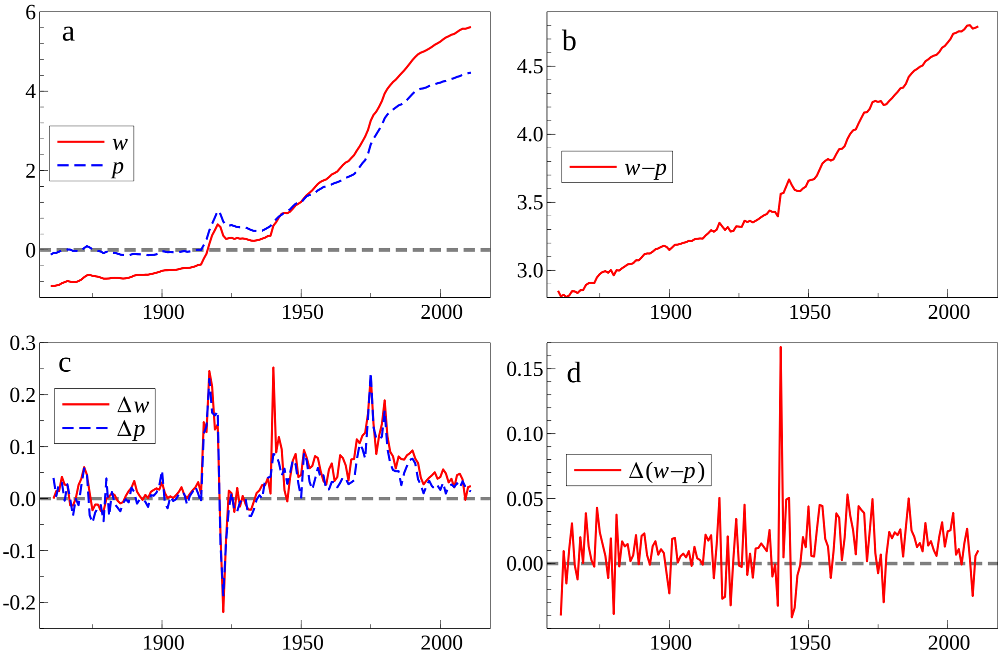

Figure 2: (a) nominal wages and prices; (b) real wages; (c) wage and price inflation; (d) real wage inflation.

The underlying theoretical relationship is that the real wage is driven by the real marginal revenue product of labour, proxied by output per worker, $(y-l)_t$. We anticipate a positive sign on labour productivity, and expect full adjustment in the long run, captured by using real unit labour costs adjusted for hours, denoted $(ulc-p)_t$, as an approximation to the equilibrium-correction mechanism (EqCM), $(w-p-y+l)_t - \widehat{\mu}$, where $\widehat{\mu}$ is the sample mean over 1860–2004 (see §9). Note that $(w-p-y+l)_t$ is also labour's share in national income. We include the unemployment rate, $U_{r,t}$, allowing for a 'Phillips curve' relationship, lowering wages when the unemployment rate is high. We also explore the possible role of a change in the unemployment rate in the dynamic modelling section where we find a tentative role for it. Finally, we include price inflation both to reflect the conditional and marginal factorization undertaken in modelling real wages, and as a 'catch-up' by workers when wages have been eroded due to less than complete adjustments to past inflation. We find the price inflation term enters non-linearly, capturing wage-price spirals.

4.2 The previous non-linear model

The non-linearity index test in §3.1 was applied to a linear model of real wage growth, where the regressors include an intercept, $\Delta(w-p)_{t-i}$ and $(ulc-p)_{t-i}$ for $i=1,2$ and $\Delta(y-l)_{t-j}$, $U_{r,t-j}$ and Δp_{t-j} for $j=0,1,2$. The test is significant at $\mathsf{p}=0.006$ with $\mathsf{F}(36,91)=1.95$. Castle and Hendry (2009) also found the index test to be significant, so we proceed to investigate non-linear models, beginning with their non-linear formulation:

$$f_t = \frac{-1}{1+1000(\Delta p_t)^2}. \tag{17}$$

The non-linear mapping in (17) is U-shaped: workers become more attentive when price inflation rises, and act to prevent further erosion of their real wages (compare the model of inattentive producers in Reis, 2006), whereas employers cut nominal wages when prices fall. Such behaviour generates wage-price spirals. Re-estimating their model on the updated data delivers similar results to those reported earlier:

$$\begin{aligned}
\Delta(w-p)_t &= \underset{(0.002)}{0.010} + \underset{(0.126)}{0.649}\,(f_t\Delta p_t) + \underset{(0.045)}{0.384}\Delta(y-l)_t + \underset{(0.048)}{0.159}\Delta(y-l)_{t-2} \\
&- \underset{(0.010)}{0.063}\,(ulc-p)_{t-2} - \underset{(0.044)}{0.129}\Delta_2 U_{r,t-1} + \underset{(0.013)}{0.029} I_{1918} + \underset{(0.013)}{0.139} I_{1940} \\
&+ \underset{(0.006)}{0.032}\,(I_{1942}+I_{1943}-I_{1944}-I_{1945}) - \underset{(0.009)}{0.041}\,(I_{1975}+I_{1977}) \qquad (18)\\
\mathsf{R}^2 &= 0.733;\ \ \widehat{\sigma} = 1.24\%;\ \ \mathit{SIC} = -5.66; \\
\chi^2_{nd}(2) &= 2.21;\ \ \mathsf{F}_{ar}(2,130) = 0.766;\ \ \mathsf{F}_{arch}(1,140) = 0.109; \\
\mathsf{F}_{het}(13,126) &= 0.794; \mathsf{F}_{reset}(2,130) = 0.106; \mathsf{F}_{chow}(7,132) = 1.354; T = 1864-2004.
\end{aligned}$$

In (18), R^2 is the squared multiple correlation, $\widehat{\sigma}$ is the residual standard deviation, coefficient standard errors are shown in parentheses and SIC is the Schwarz criterion (see Schwarz, 1978). The diagnostic tests are of the form $\mathsf{F}_{\mathsf{j}}(k, T-l)$ which denotes an approximate F-test against the alternative hypothesis j for: k^{th}-order serial correlation (F_{ar}: see Godfrey, 1978), k^{th}-order autoregressive conditional heteroskedasticity ($\mathsf{F}_{\mathsf{arch}}$: see Engle, 1982), heteroskedasticity ($\mathsf{F}_{\mathsf{het}}$: see White, 1980); the RESET test ($\mathsf{F}_{\mathsf{reset}}$: see Ramsey, 1969); parameter constancy ($\mathsf{F}_{\mathsf{Chow}}$: see Chow, 1960) over k periods; and a chi-square test for normality ($\chi^2_{\mathsf{nd}}(2)$: see Doornik and Hansen, 2008). Finally, * and ** denote significant at 5% and 1% respectively. Figure 3 records the model fit, residuals, 1-step forecasts with 95% forecast intervals, and the residual density for this baseline model.

Overall, the update is close to the original despite data revisions, and is relatively constant over the 'Great Recession'.

4.3 An approximating non-linear model

In fact, f_t is a variant of an LSTAR in π_t^2, where $\pi_t = 100\Delta p_t$ (annual inflation measured as a percentage), given by (scaling to the same mean and range as f_t):

$$Lp_t = 2\left(1+\exp(-\gamma\pi_t^2)\right)^{-1} - 2 \tag{19}$$

so the approximation in (10) becomes:

$$\alpha_1\Delta p_t + \alpha_2(\Delta p_t)^3 + \alpha_3(\Delta p_t)^4 \tag{20}$$

While investigating the polynomial approximation to such non-linearities in the wages model, we also included the most significant non-linear function of the other regressors, $U_{r,t}^2$. Selecting at 1% yields the equivalent of (11) being estimated as (adding all the other non-linear transformations of the three

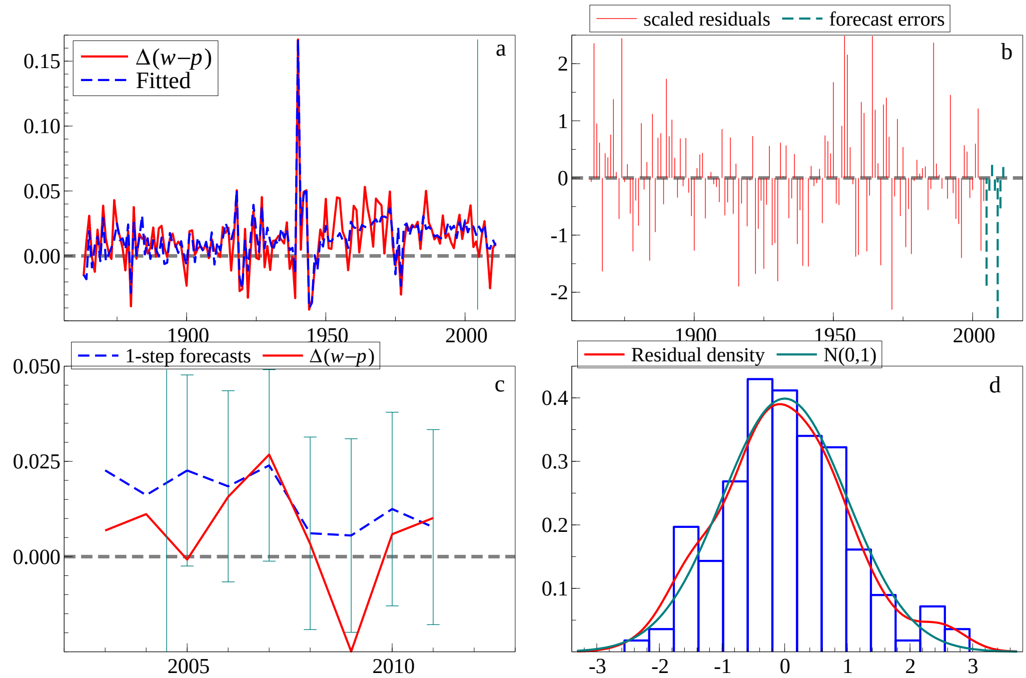

Figure 3: Equation (18): (a) actual and fitted values; (b) scaled residuals and forecast errors; (c) 1-step forecasts over the extended data, 2005–2011; (d) residual density.

explanatory variables, namely demeaned squares, cubics and exponentials, did not produce any significant improvement):

$$
\begin{aligned}
\Delta(w-p)_t = {} & \underset{(0.003)}{0.017} + \underset{(0.050)}{0.314}\Delta(y-l)_t + \underset{(0.055)}{0.184}\Delta(y-l)_{t-2} - \underset{(0.013)}{0.060}(ulc-p)_{t-2} \\
& - \underset{(0.042)}{0.166}U_{r,t} + \underset{(0.80)}{2.59}U_{r,t}^2 - \underset{(0.050)}{0.096}\Delta_2 U_{r,t} + \underset{(1.63)}{6.60}(\Delta p_t)^3 - \underset{(5.44)}{17.7}(\Delta p_t)^4 \\
& - \underset{(0.045)}{0.186}\Delta p_t - \underset{(0.03)}{0.120}\Delta^2 p_{t-1} + \underset{(0.013)}{0.148}I_{1940} - \underset{(0.013)}{0.044}I_{1944} - \underset{(0.013)}{0.052}I_{1945} - \underset{(0.013)}{0.038}I_{1977} \\
\mathsf{R}^2 = {} & 0.747; \quad \widehat{\sigma} = 1.23\%; \quad SIC = -5.55; \qquad (21) \\
\chi^2_{nd}(2) = {} & 0.88; \quad \mathsf{F}_{ar}(2,124) = 0.63; \quad \mathsf{F}_{arch}(1,139) = 0.18; \\
\mathsf{F}_{het}(19,117) = {} & 1.12; \quad \mathsf{F}_{chow}(7,126) = 1.61; \quad T = 1864 - 2004.
\end{aligned}
$$

Both non-linear terms in inflation are highly significant, and the fit and mis-specification tests are similar to (18). A comparison with the corresponding linear dynamic equation shows that three indicators have been eliminated (I_{1922}, I_{1939} and I_{1942}), in favour of the three non-linear terms ($U_{r,t}^2$, $(\Delta p_t)^3$ and $(\Delta p_t)^4$). Hence, IIS does not 'substitute' for included non-linearities when they matter. Conversely, no non-linear terms were significant when all indicators were eliminated, emphasizing their interactions.

We also added the second difference of the unemployment rate found in their earlier study: it was marginally significant but did not eliminate the need for the level or the square of $U_{r,t}$. Since $U_{r,t}$ is intrinsically positive, the combined term, $-0.18U_{r,t}(1-13.7U_{r,t})$, is negative till the unemployment rate exceeds 7.25% then becomes positive. Such an effect could represent movements along the marginal product curve, raising real wages of those still employed as employment fell, from more capital per worker and the unemployment of less productive workers. Eliasson (1999) finds a related non-linearity between unemployment and inflation in Australia, where the impact of unemployment on inflation becomes positive at higher levels of unemployment.

The non-linear inflation terms approximate the finding in Castle and Hendry (2009) of a response of real wages to inflation dependent on the level of inflation, such that low rates of inflation are apparently ignored by workers and employers, but the response rises to 1-1 at high rates. We next investigate whether an LSTAR non-linear specification eliminates the polynomial functions in (21). This illustrates our semi-automatic approach, as an automated method first selects (21), after which we try to refine this with a specific theory-driven non-linear real wage reaction to inflation.

4.4 A nesting non-linear model

First, however, adding the earlier non-linear reaction $(f_t \Delta p_t)$ to (21) makes the cubic and quadratic terms in inflation individually and jointly insignificant at 1% (but not at 5%). Equation (22) reports the estimates of the resulting model.

$$\begin{aligned} \Delta(w-p)_t &= \underset{(0.003)}{0.015} + \underset{(0.047)}{0.348}\Delta(y-l)_t + \underset{(0.053)}{0.204}\Delta(y-l)_{t-2} - \underset{(0.011)}{0.061}(ulc-p)_{t-2} \\ &\quad - \underset{(0.039)}{0.157}U_{r,t} + \underset{(0.79)}{2.56}U_{r,t}^2 - \underset{(0.050)}{0.166}\Delta_2 U_{r,t} + \underset{(0.14)}{0.625}(f_t\Delta p_t) - \underset{(0.031)}{0.131}\Delta^2 p_t \\ &\quad + \underset{(0.013)}{0.138}I_{1940} - \underset{(0.013)}{0.042}I_{1944} - \underset{(0.013)}{0.045}I_{1945} - \underset{(0.012)}{0.046}I_{1977} \qquad (22) \\ \mathsf{R}^2 &= 0.747;\;\; \widehat{\sigma} = 1.22\%;\;\; \mathit{SIC} = -5.61; \\ \chi^2_{nd}(2) &= 0.54;\;\; \mathsf{F}_{ar}(2,126) = 0.96;\;\; \mathsf{F}_{arch}(1,139) = 0.06; \\ \mathsf{F}_{het}(15,121) &= 1.26;\;\; \mathsf{F}_{reset}(2,126) = 0.28;\;\; \mathsf{F}_{chow}(7,128) = 1.09;\;\; T = 1864-2004. \end{aligned}$$

Equation (22) suggests that the wage-price spiral term is not sufficient to model all the non-linearity, but does explain the non-linear impact of inflation on real-wage growth. Some of the restricted dummies in (18) are no longer significant so are excluded. The graphs of fitted and actual values, scaled residuals and forecast errors, residual density and residual autocorrelation function are reported in Figure 4.

4.5 An LSTAR model

Replacing $f_t\Delta p_t$ in (22) by $Lp_t\Delta p_t$, non-linear estimation leads to $\widehat{\gamma} = 0.059$ as in (23). Expressed as an LSTAR model:

$$\begin{aligned} \Delta(w-p)_t &= \underset{(0.003)}{0.018} + \underset{(0.049)}{0.308}\Delta(y-l)_t + \underset{(0.053)}{0.206}\Delta(y-l)_{t-2} - \underset{(0.013)}{0.074}(ulc-p)_{t-2} \\ &\quad - \underset{(0.042)}{0.183}U_{r,t} + \underset{(0.80)}{2.64}U_{r,t}^2 - \underset{(0.049)}{0.152}\Delta_2 U_{r,t} + \underset{(0.23)}{0.822}\Delta p_t\left(1+\exp\left(-\underset{(0.023)}{0.059}\pi_t^2\right)\right)^{-1} \\ &\quad - \underset{(0.24)}{1.02}\Delta^2 p_t - \underset{(0.24)}{0.907}\Delta p_{t-1} + \underset{(0.013)}{0.140}I_{1940} - \underset{(0.013)}{0.045}I_{1944} - \underset{(0.013)}{0.048}I_{1945} - \underset{(0.013)}{0.043}I_{1977} \\ \mathsf{R}^2 &= 0.752;\;\; \widehat{\sigma} = 1.22\%;\;\; \mathit{SIC} = -5.60; \qquad (23) \\ \chi^2_{nd}(2) &= 0.31;\;\; \mathsf{F}_{ar}(2,124) = 1.26;\;\; \mathsf{F}_{arch}(1,139) = 0.14; \\ \mathsf{F}_{het}(15,121) &= 1.81^*;\;\; \mathsf{F}_{chow}(7,126) = 1.31;\;\; T = 1864-2004. \end{aligned}$$

Then Lp_t generates almost identical behaviour to f_t, as seen in Figure 5. The two series have a correlation of 0.96, but Lp_t rises more steeply around the origin, so would generate a faster wage-price spiral as inflation rose. However, neither model (23) and (21) encompasses the other as Table 8 shows.

Similarly, neither $Lp_t\Delta p$ is significant if added to (22), nor $f_t\Delta p_t$ when added to (23), so they too are close substitutes. Consequently, some of the considerations in section 2 may apply, although an important difference is that the transition is exogenous here, as against the lagged dependent variable

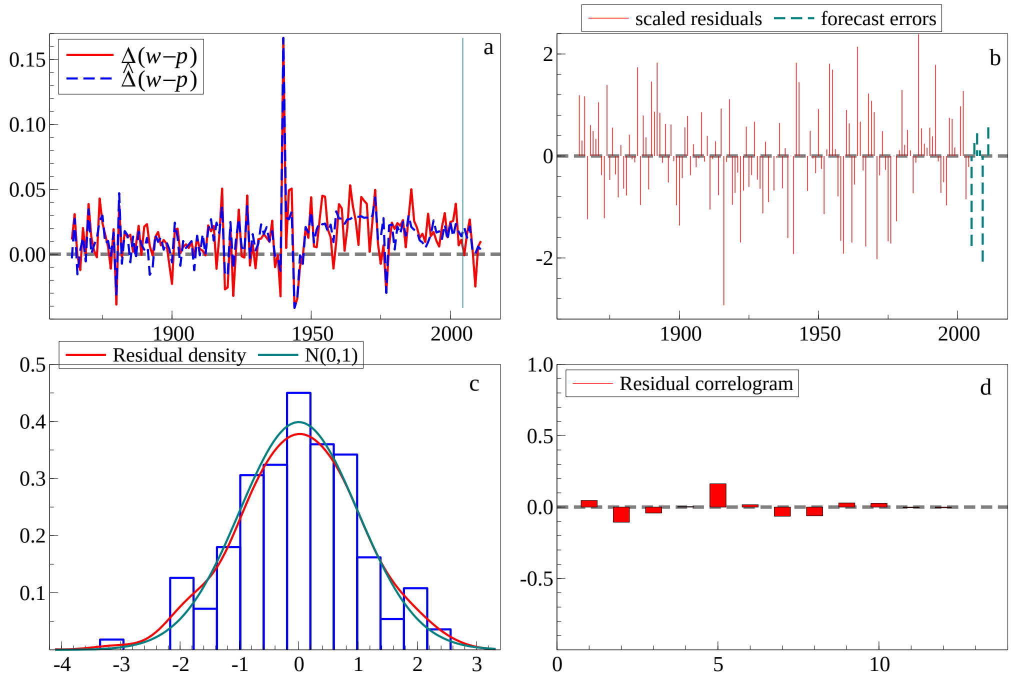

Figure 4: Equation (22): (a) actual and fitted values; (b) scaled residuals and forecast errors; (c) residual density; (d) residual autocorrelation function.

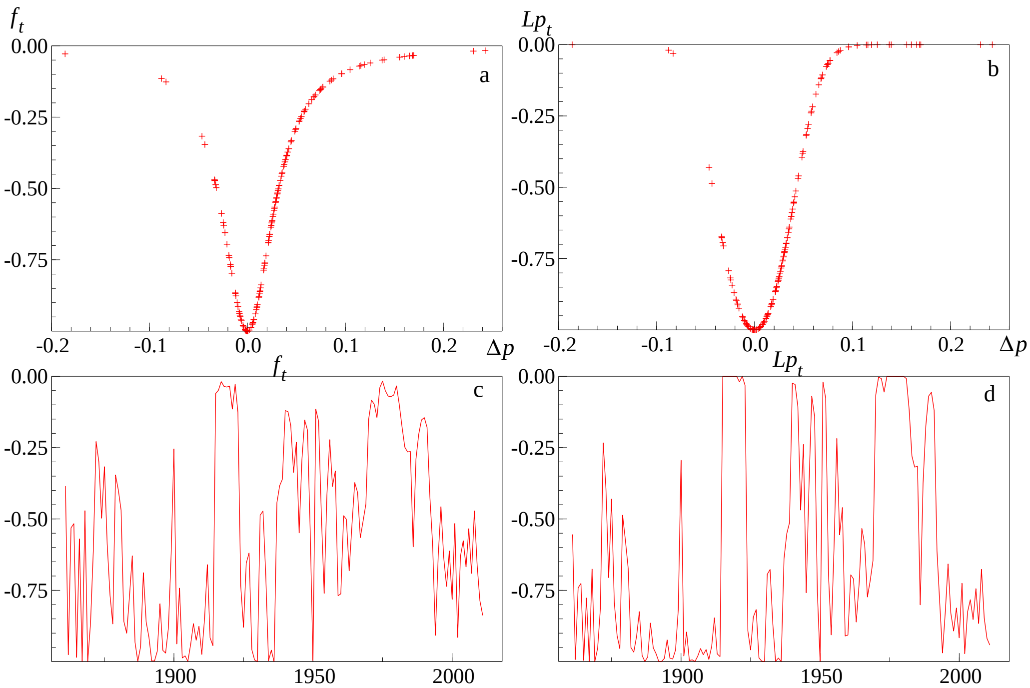

Figure 5: (a) f_t against Δp_t; (b) Lp_t against Δp_t; (c) f_t; (d) Lp_t.

earlier. Reparameterizing the model as an LSTAR yields apparently odd looking coefficients on the inflation variables, but if (19) replaces the LSTAR term, only $\Delta^2 p_t$ remains and is close to that in (22). Thus, despite Δp_t 'entering' $(\Delta w_t - \Delta p_t)$, real wages are primarily determined by forces different from

Test	Model 1 vs. Model 2	Model 2 vs. Model 1
Cox N(0,1)	-4.66^{**}	-7.25^{**}
Joint Model	F(2,125) = 4.08*	F(1,125) = 10.9**

Table 8: Encompassing tests of (21) against (23).

nominal prices, consistent with the 'Classical dichotomy': in particular, the impact of Δp_t on real wages is zero at high inflation.

4.6 An alternative non-linear model

A further alternative non-linear specification is that reported by Nielsen (2009) (general model, column D), reported in equation (24). Results are similar to those reported in his paper.[5]

$$\begin{aligned}
\Delta(w-p)_t &= \underset{(0.002)}{0.006} + \underset{(0.126)}{0.882}(f_t\Delta p_t) + \underset{(0.045)}{0.297}\Delta(y-l)_t - \underset{(0.013)}{0.072}(ulc-p)_{t-2} \\
&\quad -\underset{(0.045)}{0.148}\Delta_2 U_{r,t} + \underset{(0.0001)}{0.0003}\left(I_{1860-1913}\times \Delta U_r^{-1}\right)_t \\
&\quad +\underset{(0.00006)}{0.0003}\left(I_{1860-1913}\times U_r^{-1}\right)_{t-2} - \underset{(0.008)}{0.031}\left(I_{1947-2011}\times \Delta log\left(U_r\right)\right)_t \\
&\quad -\underset{(0.0009)}{0.004}\left(I_{1947-2011}\times log\left(U_r\right)\right)_{t-1} + \underset{(0.012)}{0.036}I_{1918} + \underset{(0.012)}{0.146}I_{1940} \\
&\quad +\underset{(0.006)}{0.039}\left(I_{1942}+I_{1943}-I_{1944}-I_{1945}\right) - \underset{(0.008)}{0.037}\left(I_{1975}+I_{1977}\right) \qquad (24) \\
\mathsf{R}^2 &= 0.783;\ \widehat{\sigma} = 1.13\%;\ SIC = -5.77; \\
\chi^2_{nd}(2) &= 0.53;\ \mathsf{F}_{ar}(2,126) = 0.089;\ \mathsf{F}_{arch}(1,139) = 0.003; \\
\mathsf{F}_{het}(19,119) &= 1.093; \mathsf{F}_{reset}(2,126) = 2.228; \mathsf{F}_{chow}(7,128) = 1.218; T = 1864-2004.
\end{aligned}$$

The regime-shift variables matter, and indeed remain relevant over the 'Great Recession', as curtailing their influence to end in 2004 leads to a marked deterioration in RMSFE.

4.7 Encompassing

To shed light on which models may be preferred, we examine a range of pairwise encompassing tests. For comparison, we estimate the nesting model which has 24 parameters, denoted 'Nest' in table 9. This model has an equation standard error of 1.01%, log-likelihood of 461.97 and $SIC = -5.68$. The non-linear models reported above are then tested for encompassing and results are reported in table 9: combined dummies are entered separately for estimation, but the encompassing tests are computed over regressors other than dummies. The third column reports encompassing tests of the non-linear models against the nesting model, where the additional variables in the nesting model are tested for their significance. The following columns undertake pairwise encompassing tests; for example, the column for (23) against row (18) tests for the additional regressors in (18) compared to (23).

The table demonstrates that no model dominates on an encompassing citerion. Hence, the non-linearities in the form of polynomials, smooth transitions and regime shifts can all approximate the non-linear reaction of real wages to inflation over the century and a half examined, yet none captures all the effects.

[5] Data revisions to the extended dataset result in slightly different estimates.

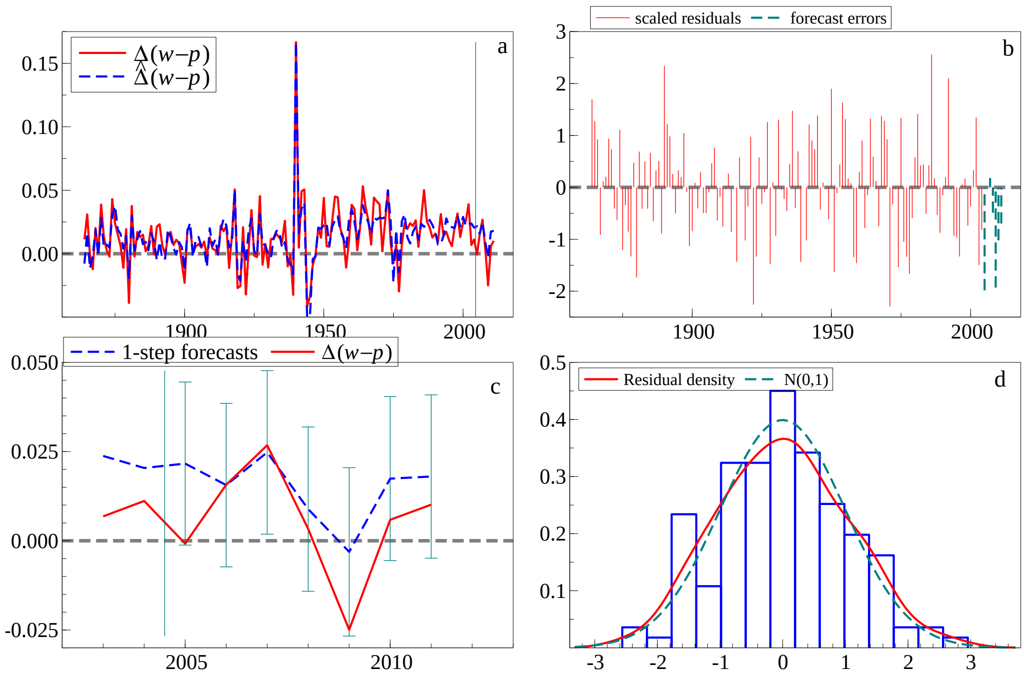

Figure 6: Equation (24): (a) actual and fitted values; (b) scaled residuals and forecast errors; (c) 1-step forecasts over the extended data, 2005–2011; (d) residual density

	$\widehat{l}$	Nest	(18)	(22)	(23)	(24)
Nest	462.0					
(18)	423.8	$\chi^2(11) = 76.4^{**}$			$F(1, 121) = 5.52^{*}$	$F(1, 127) = 7.83^{**}$
(22)	428.0	$\chi^2(12) = 68.0^{**}$	$F(3, 128) = 12.54^{**}$		$F(1, 117) = 8.60^{**}$	$F(4, 124) = 4.27^{**}$
(23)	429.5	$\chi^2(10) = 64.9^{**}$	$F(6, 121) = 4.22^{**}$	$F(2, 117) = 3.46^{*}$		$F(6, 117) = 3.63^{**}$
(24)	439.1	$\chi^2(8) = 45.7^{**}$	$F(4, 127) = 10.98^{**}$	$F(4, 124) = 6.04^{**}$	$F(5, 116) = 7.00^{**}$	

Table 9: Encompassing tests for non-linear models. $\widehat{l}$ is the log-likelihood.

5 A step-indicator saturation equation

Doornik *et al.* (2013) propose a generalization of IIS using step-indicator saturation (SIS), adding a complete set of step indicators $\mathcal{S}_1 = \left\{1_{\{t \leq j\}}, j = 1, \ldots, T\right\}$, where $1_{\{t \leq j\}} = 1$ for observations up to j, and zero otherwise. Step indicators are the cumulation of impulse indicators up to each next observation illustrated as follows:

$$\begin{array}{cc} \text{IIS: impulse-indicators} & \text{SIS: step-indicators} \\ \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & \ddots \end{bmatrix} & \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

SIS has the correct null retention frequency of α in constant conditional models for a nominal test size of α. The approximate alternative retention-frequency function has been derived analytically by Doornik *et al.* (2013) for simple models, and shows much higher probabilities of retaining location shifts than IIS, yet a similar potency for impulses by 2 successive equal-magnitude opposite-signed steps. Two successive opposite-signed steps of different magnitudes capture both a location shift and an impulse.

To check the robustness of the earlier models, we applied SIS (now available in *Autometrics*) to a GUM which also nested both (22) (so implicitly (23) as well) and (24), and found a substantively improved representation, in which $(w-p-y+l)$ replaced the measure $(ulc-p)$ that was adjusted for changes in hours. This is reported in (25) where $\widehat{\mu}$ is the sample mean of $(w-p-y+l)$, $u_{r,t}=\log(U_{r,t})$, and (e.g.) S_{1939} is the step indicator that is unity till 1939 and zero thereafter.

$$
\begin{aligned}
\Delta\left(w-p\right)_{t} = {} & \underset{(0.003)}{0.030} + \underset{(0.042)}{0.354}\Delta\left(y-l\right)_{t} + \underset{(0.034)}{0.116}\Delta_{2}\left(y-l\right)_{t-1} - \underset{(0.028)}{0.179}\left(w-p-y+l-\widehat{\mu}\right)_{t-2} \\
& - \underset{(0.034)}{0.178}U_{r,t} + \underset{(0.68)}{2.68}U_{r,t}^{2} - \underset{(0.045)}{0.13}\Delta_{2}U_{r,t} + \underset{(0.012)}{0.711}\left(f_{t}\Delta p_{t}\right) - \underset{(0.029)}{0.130}\Delta^{2}p_{t-1} \\
& - \underset{(0.011)}{0.145}S_{1939} + \underset{(0.015)}{0.176}S_{1940} - \underset{(0.011)}{0.058}S_{1941} - \underset{(0.008)}{0.024}(S_{2011}-S_{1946})\Delta u_{r,t} \\
& - \underset{(0.011)}{0.036}I_{1916} + \underset{(0.006)}{0.027}\left(I_{1942}+I_{1943}-I_{1944}-I_{1945}\right) - \underset{(0.011)}{0.044}I_{1977} \qquad (25)
\end{aligned}
$$

$$
\mathsf{R}^{2}=0.820;\quad \widehat{\sigma}=1.04\%;\quad SIC=-5.85;
$$

$$
\chi^{2}_{nd}\left(2\right)=2.26;\quad \mathsf{F}_{ar}\left(2,123\right)=0.39;\quad \mathsf{F}_{arch}\left(1,139\right)=0.49;\quad \mathsf{F}_{reset}\left(2,124\right)=2.28;
$$

$$
\mathsf{F}_{het}\left(20,116\right)=0.82;\quad \mathsf{F}_{chow}\left(7,125\right)=0.95;\quad T=1864-2004.
$$

By design, (25) encompasses the previous models, and all the mis-specification tests are insignificant. This model reveals that most of the variables in common with (22) have similar coefficients, other than a stronger and more rapid feedback of almost -0.18 from the previous labour share, and replacing $\Delta\left(y-l\right)_{t-2}$ by $\Delta_{2}\left(y-l\right)_{t-1}$, as well as switching from pure impulse dummies to a mixture of steps and impulses. Two of the variables from (24) are also retained, so an interaction of a step shift with a variable matters as well. However, the main role of the step indicators seems to be explaining the much higher average growth rate of real wages post war (1.8% p.a., versus 0.7% p.a. pre-1945), even though $\Delta(y-l)$ is included and displays a similar pattern. Figure 7 reports the graphical statistics.

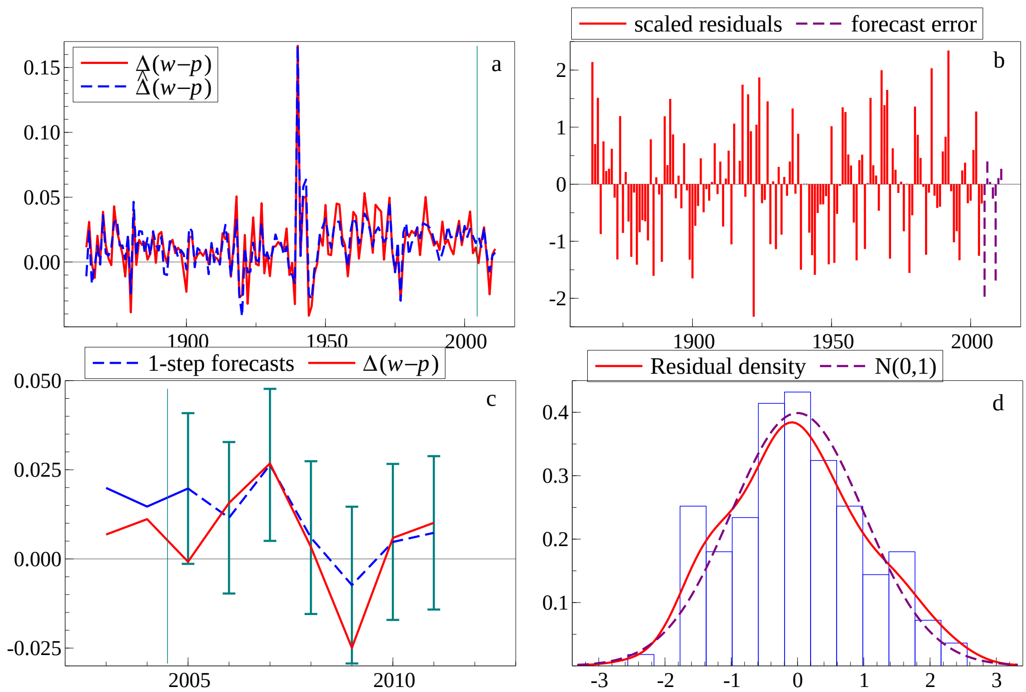

Figure 7: Equation (25): (a) actual and fitted values; (b) scaled residuals and forecast errors; (c) 1-step forecasts; (d) residual density

6 Testing exogeneity

IIS can be used to test the exogeneity of the conditioning variables as in Hendry and Santos (2010). Under the null of super exogeneity, the parameters in the conditional model are invariant to shifts in the marginal models, so indicators in the latter should not enter the former. A vector autoregression (VAR) in the system of four variables $(w-p)_t$, $(y-l)_t$, Δp_t and $U_{r,t}$ was selected with IIS, and the additional impulse indicators in the three marginal models were then tested for significance in (22). The same procedure using SIS on the 3 marginal VAR equations was applied to (25). Table 10 reports the results.

Variable	null distribution	IIS test statistic	null distribution	SIS test statistic
$(y-l)_t$	F(11,117)	1.16	F(2,123)	0.77
Δp_t	F(11,117)	1.22	F(7,118)	1.87
$U_{r,t}$	F(9,118)	1.05	F(14,111)	1.37
Joint	F(16,112)	1.22	F(20,105)	1.41

Table 10: IIS super-exogeneity tests of (22) and SIS tests of (25).

While none of the tests rejects exogeneity, there are 3 impulse indicators in common between (22) and the marginal equation for $(y-l)_t$, namely $I_{1940},\ I_{1944},\ I_{1945}$, although their values would not be consistent with only entering $\Delta(w-p)_t$ through $(y-l)_t$. I_{1940} is also in common with the equation for Δp_t, but is positive and at a much smaller value: the spike in real wages engineered at the start of the Second World War was a 'separate' event. There are no step indicators in common with (25) when selecting in each of the VAR equations at $\alpha = 0.001$, although 20 separate step indicators are retained across the three marginal models.

7 Forecasting

Ex post forecasts for $\Delta(w-p)_t$ have previously been shown graphically for several of the models above. Here, Figure 8 records the 1-step ahead forecasts from parameter estimates over 1864–2004, for the levels of real wages, with and without intercept correction (IC) for the non-linear models reported in §4.4, §4.6, and §5. The IC used was the average residual over 2003–2004, and the 95% forecast intervals shown by error bars in Figure 8 allow for parameter uncertainty. All these forecasts exhibit similar patterns. Table 11 reports the RMSFEs of these three non-linear specifications for $\Delta(w-p)_t$, as well as a random-walk model of $(w-p)_t$ (RW), and the forecasts from the VAR, both with IIS at 1%, with and without intercept corrections for the real-wage equation (1-step RMSFEs are the same for levels' forecasts).

The original forecasts tend to miss the downturn in 2009, though (25) comes close, but all the IC forecasts do well and are relatively similar. Although all the non-linear models somewhat outperform the linear, they have contemporaneous regressors, albeit exogenous. Overall, the SIS encompassing model has the smallest RMSFEs, especially in equations without ICs, and is close to its in-sample $\widehat{\sigma}$. This is an unusual result because (25) is a complicated non-linear specification. The forecasting literature often finds that the forecast performance of non-linear models is not good in comparison to linear models (see, for example, De Gooijer and Kumar, 1992, Clements and Smith, 1999, and Clements, Franses and Swanson, 2004), and even more so when facing breaks (see Castle *et al.*, 2011).

8 Conclusions

The empirical study confirmed the need for joint modelling of dynamics, location shifts, relevant variables and non-linearities. Failing to include any of these features led to substantive mis-specifications, with

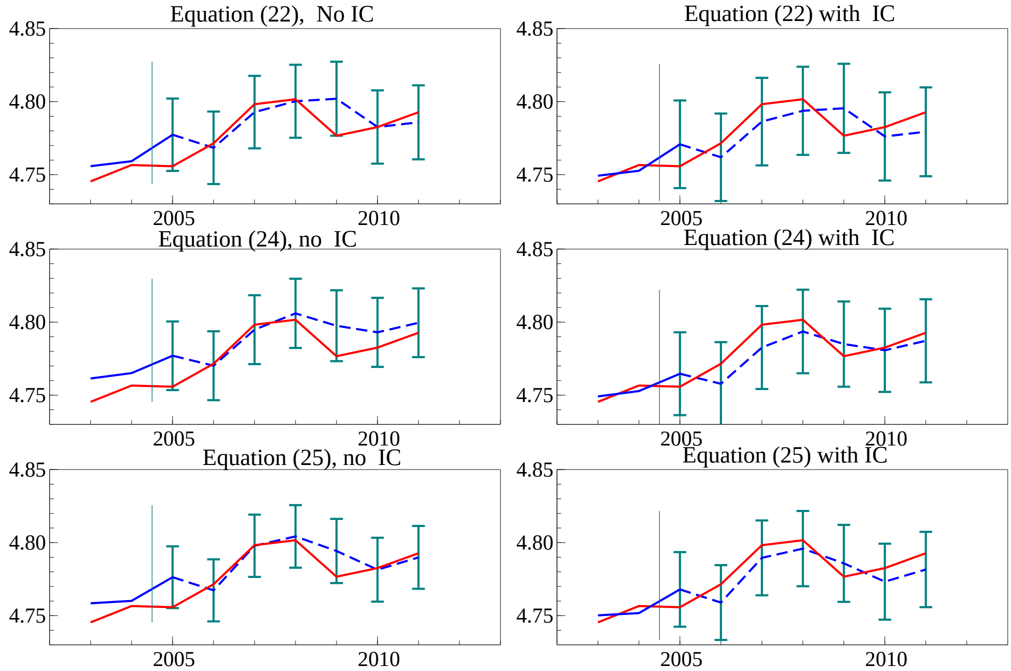

Figure 8: Forecasts for real wages with and without intercept corrections. The top row records equation (22), the second row, (24) and the bottom row (25).

Equation	$\widehat{\sigma}$	No IC	IC
(22)	1.22%	1.31%	1.25%
(24)	1.13%	1.23%	1.04%
(25)	1.04%	1.05%	1.00%
RW	2.23%	1.57%	1.54%
VAR	1.67%	2.37%	1.54%

Table 11: RMSFEs of forecasts of $\Delta(w-p)_{T+h}$ with and without intercept corrections, with in-sample equation standard error for comparison.

included variables being insignificant in restricted formulations, yet important in more general models. Automatic model selection seems a viable approach to tackling all the complications jointly, even when there are more candidate variables than observations.

There are three important economic implications. First, there is a wage-price spiral of increasing reactions of real wages to inflation as inflation rises in absolute terms. That adds persistence to the wage-price process, and may be what creates the impression of 'sticky inflation'. Such a non-linear adjustment can be approximated in several ways, and doubtless there are other ways than those considered above. Second, real wages are primarily determined by forces different from nominal prices, consistent with the 'Classical dichotomy'. Third, using a general polynomial led to an additional non-linearity in unemployment, which suggested that real wages rise with unemployment beyond about 7.25%, probably from rising marginal productivity, rather than wage bargaining. That finding is consistent with the presence of involuntary unemployment, as no evidence of any reverse relation of high real wages causing unemployment was found in Hendry (2001).

The Monte Carlo simulations of TAR and LSTAR models showed the difficulty of detecting and estimating regime shifts. Despite that difficulty, the empirical evidence for non-linear adjustments of real wages to inflation is clear cut. Basing the reaction on an exogenous variable seems to explain that difference. Moreover, applying either impulse-indicator saturation or step-indicator saturation did not preclude finding non-linearites, nor did those modelled non-linearities obviate the need for including a number of indicators. Conversely, not removing large outliers or shifts could hide the presence of other variables, including non-linearities.

The forecasting results over the 'Great Recession' rebut the notion that parsimony is essential, as the most complicated model produced the smallest 1-step RMSFEs. However, almost all methods benefitted from intercept corrections setting their forecasts 'back on track' at the forecast origin.

References

Attfield, C. L. F., Demery, D., and Duck, N. W. (1995). Estimating the UK demand for money function: A test of two approaches. Mimeo, Economics Department, University of Bristol.

Castle, J. L. (2005). Evaluating PcGets and RETINA as automatic model selection algorithms. *Oxford Bulletin of Economics and Statistics*, **67**, 837–880.

Castle, J. L., Doornik, J. A., and Hendry, D. F. (2011). Evaluating automatic model selection. *Journal of Time Series Econometrics*, **3 (1)**, DOI: 10.2202/1941–1928.1097.

Castle, J. L., Doornik, J. A., and Hendry, D. F. (2012). Model selection when there are multiple breaks. *Journal of Econometrics*, **169**, 239–246.

Castle, J. L., Doornik, J. A., Hendry, D. F., and Nymoen, R. (2012b). Mis-specification testing: Non-invariance of expectations models of inflation. *Econometric Reviews*, forthcoming.

Castle, J. L., Fawcett, N. W. P., and Hendry, D. F. (2011). Forecasting Breaks and During Breaks. In Clements, M. P., and Hendry, D. F. (eds.), *Oxford Handbook of Economic Forecasting*, pp. 315–353. Oxford: Oxford University Press.

Castle, J. L., and Hendry, D. F. (2009). The long-run determinants of UK wages, 1860–2004. *Journal of Macroeconomics*, **31**, 5–28.

Castle, J. L., and Hendry, D. F. (2010). A low-dimension portmanteau test for non-linearity. *Journal of Econometrics*, **158(2)**, 231–245.

Castle, J. L., and Hendry, D. F. (2011). Automatic selection of non-linear models. In Wang, L., Garnier, H., and Jackman, T. (eds.), *System Identification, Environmental Modelling and Control*, pp. 229–250. New York: Springer.

Castle, J. L., and Shephard, N. (eds.)(2009). *The Methodology and Practice of Econometrics*. Oxford: Oxford University Press.

Chan, K. S., and Tong, H. (1986). On estimating thresholds in autoregressive models. *Journal of Time Series Analysis*, **7**, 179–194.

Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica*, **28**, 591–605.

Clements, M. P., Franses, P. F., and Swanson, N. (2004). Forecasting economic and financial time-series with non-linear models. *International Journal of Forecasting*, **20**, 169–183.

Clements, M. P., and Krolzig, H.-M. (1998). A comparison of the forecast performance of Markov-switching and threshold autoregressive models of US GNP. *Econometrics Journal*, **1**, C47–C75.

Clements, M. P., and Smith, J. (1999). A Monte Carlo study of the forecasting performance of empirical SETAR models. *Journal of Applied Econometrics*, **14**, 124–141.

Crafts, N. F. R., and Mills, T. C. (1994). Trends in real wages in Britain, 1750-1913. *Explorations in Economic History*, **31**, 176–194.

De Gooijer, J. G., and Kumar, K. (1992). Some recent developments in non-linear time series modelling, testing and forecasting. *International Journal of Forecasting*, **8**, 135–156.

Dicks-Mireaux, L. A., and Dow, J. C. R. (1959). The determinants of wage inflation: United Kingdom, 1946–1956. *Journal of the Royal Statistical Society*, **A, 122**, 145–84.

Doornik, J. A. (2009a). Autometrics. In Castle, and Shephard (2009), pp. 88–121.

Doornik, J. A. (2009b). Econometric model selection with more variables than observations. Working paper, Economics Department, University of Oxford.

Doornik, J. A., Hendry, D. F., and Pretis, F. (2013). Step-indicator saturation. Working paper, Economics Department, Oxford University.

Doornik, J. A., and Hansen, H. (2008). An omnibus test for univariate and multivariate normality. *Oxford Bulletin of Economics and Statistics*, **70**, 927–939.

Eliasson, A.-C. (1999). *Smooth Transitions in Macroeconomic Relationships*. Stockholm: Economic Research Institute, Stockholm School of Economics: ISBN: 91-7258-516-1.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity, with estimates of the variance of United Kingdom inflation. *Econometrica*, **50**, 987–1007.

Ericsson, N. R., Hendry, D. F., and Prestwich, K. M. (1998). The demand for broad money in the United Kingdom, 1878–1993. *Scandinavian Journal of Economics*, **100**, 289–324.

Feinstein, C. H. (1972). *National Income, Expenditure and Output of the United Kingdom, 1855–1965*. Cambridge: Cambridge University Press.

Feinstein, C. H. (1990). New estimates of average earnings in the UK, 1880-1913. *Economic History Review*, **43**, 595–632.

Franses, P. H., and Van Dijk, D. (2000). *Non-linear Time Series Models in Empirical Finance*. Cambridge: Cambridge University Press.

Friedman, M., and Schwartz, A. J. (1982). *Monetary Trends in the United States and the United Kingdom: Their Relation to Income, Prices, and Interest Rates, 1867–1975*. Chicago: University of Chicago Press.

Galí, J., and Gertler, M. (1999). Inflation dynamics: A structural econometric analysis. *Journal of Monetary Economics*, **44**, 195–222.

Galí, J., Gertler, M., and Lopez-Salido, J. D. (2001). European inflation dynamics. *European Economic Review*, **45**, 1237–1270.

Godfrey, L. G. (1978). Testing for higher order serial correlation in regression equations when the regressors include lagged dependent variables. *Econometrica*, **46**, 1303–1313.

Godley, W. A. H., and Nordhaus, W. D. (1972). Pricing in the trade cycle. *Economic Journal*, **82**, 853–882.

Granger, C. W. J., and Teräsvirta, T. (1993). *Modelling Nonlinear Economic Relationships*. Oxford: Oxford University Press.

Hendry, D. F. (2001). Modelling UK inflation, 1875–1991. *Journal of Applied Econometrics*, **16**, 255–275.

Hendry, D. F. (2000). Epilogue: The success of general-to-specific model selection. In *Econometrics: Alchemy or Science?*, pp. 467–490. Oxford: Oxford University Press. New Edition.

Hendry, D. F. (2009). The methodology of empirical econometric modeling: Applied econometrics through the looking-glass. In Mills, T. C., and Patterson, K. D. (eds.), *Palgrave Handbook of Econometrics*, pp. 3–67. Basingstoke: Palgrave MacMillan.

Hendry, D. F., and Ericsson, N. R. (1991). An econometric analysis of UK money demand in 'Monetary

Trends in the United States and the United Kingdom' by Milton Friedman and Anna J. Schwartz. *American Economic Review*, **81**, 8–38.

Hendry, D. F., and Johansen, S. (2013). Model discovery and Trygve Haavelmo's legacy. *Econometric Theory*, forthcoming.

Hendry, D. F., Johansen, S., and Santos, C. (2008). Automatic selection of indicators in a fully saturated regression. *Computational Statistics*, **33**, 317–335. Erratum, 337–339.

Hendry, D. F., and Krolzig, H.-M. (2005). The properties of automatic Gets modelling. *Economic Journal*, **115**, C32–C61.

Hendry, D. F., and Mizon, G. E. (1998). Exogeneity, causality, and co-breaking in economic policy analysis of a small econometric model of money in the UK. *Empirical Economics*, **23**, 267–294.

Hendry, D. F., and Santos, C. (2010). An automatic test of super exogeneity. In Watson, M. W., Bollerslev, T., and Russell, J. (eds.), *Volatility and Time Series Econometrics*, pp. 164–193. Oxford: Oxford University Press.

Henry, S. (1982). *Empirical Models of Real Wages with Applications to the UK.* Discussion paper: National Institute of Economic and Social Research.

Johansen, S., and Nielsen, B. (2009). An analysis of the indicator saturation estimator as a robust regression estimator. in Castle, and Shephard (2009), pp. 1–36.

Kock, A. B., and Teräsvirta, T. (2011). Forecasting with nonlinear time series models. In Clements, M. P., and Hendry, D. F. (eds.), *Oxford Handbook of Economic Forecasting*, pp. 61–88. Oxford: Oxford University Press.

Layard, R., Nickell, S. J., and Jackman, R. (1991). *Unemployment, Macroeconomic Performance and the Labour Market*. Oxford: Oxford University Press.

Lipsey, R. G. (1960). The relationship between unemployment and the rate of change of money wage rates in the UK, 1862–1957: A further analysis. *Economica*, **27(105)**, 1–32.

Lundbergh, S., and Teräsvirta, T. (2002). Forecasting with smooth transition autoregressive models. In Clements, M. P., and Hendry, D. F. (eds.), *A Companion to Economic Forecasting*, pp. 485–509. Oxford: Blackwells.

Luukkonen, R., Saikkonen, P., and Teräsvirta, T. (1988). Testing linearity in univariate time series models. *Scandinavian Journal of Statistics*, **15**, 161–175.

Maddala, G. S. (1977). *Econometrics*. New York: McGraw-Hill.

Mitchell, B. R. (1988). *British Historical Statistics*. Cambridge: Cambridge University Press.

Nickell, S. J. (1990). Inflation and the UK labour market. *Oxford Review of Economic Policy*, **6**, 26–35.

Nielsen, H. B. (2009). Comment on 'the long-run determinants of UK wages, 1860–2004'. *Journal of Macroeconomics*, **31**, 29–34.

Perez-Amaral, T., Gallo, G. M., and White, H. (2003). A flexible tool for model building: the relevant transformation of the inputs network approach (RETINA). *Oxford Bulletin of Economics and Statistics*, **65**, 821–838.

Perez-Amaral, T., Gallo, G. M., and White, H. (2005). A comparison of complementary automatic modelling methods: RETINA and PcGets. *Econometric Theory*, **21**, 262–277.

Phillips, A. W. H. (1958). The relation between unemployment and the rate of change of money wage rates in the United Kingdom, 1861–1957. *Economica*, **25**, 283–299.

Priestley, M. B. (1981). *Spectral Analysis and Time Series*. New York: Academic Press.

Quandt, R. E. (1983). Computational problems and methods. In *Computational Problems and Methods*, pp. 699–746. Amsterdam: North Holland.

Ramsey, J. B. (1969). Tests for specification errors in classical linear least squares regression analysis. *Journal of the Royal Statistical Society B*, **31**, 350–371.

Reis, R. (2006). Inattentive producers. *Review of Economic Studies*, **73**, 793–821.

Sargan, J. D. (1964). Wages and prices in the United Kingdom: A study in econometric methodology (with discussion). In Hart, P. E., Mills, G., and Whitaker, J. K. (eds.), *Econometric Analysis for National Economic Planning*, Vol. 16 of *Colston Papers*, pp. 25–63. London: Butterworth Co.

Sargan, J. D. (1980). A model of wage-price inflation. *Review of Economic Studies*, **47**, 979–1012.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, **6**, 461–464.

Shadman-Mehta, F. (1995). An empirical study of the determinants of real wages and employment: The Phillips curve revisited. Unpublished thesis, Université Catholique de Louvain, Belgium.

Sleeman, A. (1981). The relation between unemployment and the rate of change of money wage rates in the United Kingdom, 1851-1979. Paper presented to the Atlantic Economic Society, LSE, London.

Teräsvirta, T. (1994). Specification, estimation and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association*, **89**, 208–218.

Thomas, J. J. (1984). Wages and prices in the United Kingdom, 1862-1913: A re-examination of the Phillips curve. Presentation, ESRC Quantitative Economic History Study Group, Oxford.

Tong, H. (1983). *Threshold Models in Non-Linear Time Series Analysis*. New York: Springer-Verlag.

White, H. (1980). A heteroskedastic-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, **48**, 817–838.

9 Appendix: Data definitions

Y_t	=	real GDP, £million, 1985 prices	[6], p.836, [9]a (1993), [20] code:YBHH at 2005 prices.
P_t	=	implicit deflator of GDP, (1985=1)	[6], p.836, [9]a (1993), [20] code:ABML.
U_t	=	unemployment	[7], [9]c (1993), [19] code: MGSC.
$Wpop_t$	=	working population	[7], [9]c (1993), [19] code: MGSF.
$U_{r,t}$	=	$U_t/Wpop_t$ (unemployment rate, fraction)	
L_t	=	employment (= $Wpop_t - U_t$)	[4], [5]
W_t	=	average weekly wage earnings	[17], [18], [19] code: LNMM
$W_{r,t}$	=	nominal wage rates	[5], [12], [18]
H_t	=	normal hours (from 1920)	[6], p.148, [9]
ULC_t	=	unit labour costs $(= L_tW_{r,t}/Y_t)$	
Δx_t	=	$(x_t - x_{t-1})$ for any variable x_t	
$\Delta^2 x_t$	=	$\Delta x_t - \Delta x_{t-1}$	

Sources: [1] Friedman and Schwartz (1982); [2] Attfield, Demery and Duck (1995); [3] Ericsson, Hendry and Prestwich (1998); [4] Shadman-Mehta (1995) (who cites Sleeman (1981) and Thomas (1984) as sources); [5] Phillips (1958); [6] Mitchell (1988); [7] Feinstein (1972); [8] Bank of England; [9] Bean (taken from (a) *Economic Trends Annual Supplements*, (b) *Annual Abstract of Statistics*, (c) *Department of Employment Gazette* and (d) *National Income and Expenditure*, as well as other sources cited here); [10] Cameron and Muellbauer; [11] UN Statistical Yearbook; [12] Office for National Statistics, Blue Book; [13] Board of Trade (1860–1908); [14] SS Stats; [15] Annual Abstract of Statistics; [16] Office for National Statistics, Labour Market Trends; [17] Crafts and Mills (1994); [18] Feinstein (1990); [19] Office for National Statistics, Labour Force Survey; [20] Office for National Statistics, Economic Trends Annual Supplement.

Notes:

Hendry and Ericsson (1991) and Hendry (2001) provide detailed discussions about most of these series.

Average weekly wages: a measure of full-time weekly earnings for all blue collar workers, where the coverage has been extended to include more occupations, and allows for factors such as changes in the composition of the manual labour force by age, sex, and skill, and the effect of variations in remuneration under piece rates and other systems of payments, but not adjusted for time lost through part-time work, short-time, unemployment etc. A reduction in standard hours worked that was offset by a rise in hourly wage rates would not be reflected in the index. From 1855–1880, the data are from Feinstein (1990), but not revised to increase coverage. Prior to that, the data come from a number of sources on average wage rates for blue collar workers.

Nominal wage rates: hourly wage rates prior to 1946, then weekly wage rates afterwards, so the latter were standardized by dividing by normal hours. The trend rate of decline of hours is about 0.5% p.a. (based on a drop from 56 to 40 hours per week between 1913 and 1990, with an additional increase in paid holidays), so unit labour costs were adjusted accordingly, and spliced to an average earnings index for the whole economy including bonuses [ONS: LNMM] from 1991 and rebased to 2000=1. Average earnings index discontinued in 2010, and replaced with average weekly earnings.