6

The Polymath: Combining Theory and Data

Abstract There are numerous possible approaches to building a model of a given data set, whether it be time series, cross section or panel. In economics, imposing a 'theory model' on the data, by simply estimating its parameters, is common. In 'big data' analyses, various methods of selecting relationships are used (aka 'data mining'), but in practice, modellers often select equations from data using theory-based guidelines. We discuss an approach that can retain all available theory information unaffected by selecting over additional candidate variables, lags (for time series), and non-linear functions, taking account of both potential outliers and shifts, yet can deliver an improved model when the theory specification is incomplete, incorrect, or changes over time.

6.1   Theory Driven and Data Driven Models

Two main approaches to empirically modelling a relationship are purely theory driven and purely data driven. In the former, common in economics, the putative relation is derived from a theoretical analysis claimed to represent the relevant situation, then its parameters are estimated by imposing that 'theory model' on the data.

Theory Driven Modelling
Let y_t denote the variable to be modelled by a set of n explanatory variables $\mathbf{z}_t$ when the theory relation is $y_t = f(\mathbf{z}_t)$,
then the parameters of the known function $f(\cdot)$ are estimated from a sample of data over $t = 1, \dots, T$.

In what follows, we will use a simple aggregate example based on the theory-model that monetary expansion causes inflation, reflecting Friedman's claim: 'inflation is always and everywhere a monetary phenomenon'. While it is certainly true that sufficiently large money growth can cause inflation (as in the Hungarian hyperinflation of 1945–1946), it need not do so, as the vast increase in the US monetary base from Quantitative Easing has shown, with the Federal Reserve System balances expanding by several \$trillion. Thus, our dependent variable (y_t) is the rate of inflation, related by a linear function ($f(\cdot)$), in the simplest setting to the growth rate of the money stock together with lagged values of inflation and money growth ($\mathbf{z}_t$) to reflect non-instantaneous adjustments. Previous research has established that 'narrow money' (technically called M1 for currency in circulation plus chequing accounts) does not cause inflation in the UK, so instead we consider the growth in 'broad money' (technically M4, comprising all bank deposits, although the long-run series used here is spliced *ex post* from M2, M3 and M4 as the financial system and measurements evolved over time).

In a data-driven approach, observations on a larger set of $N > n$ variables (denoted $\{\mathbf{x}_t\}$) are collected to 'explain' y_t, which here could augment money with interest rates, growth in GDP and the National Debt, excess demand for goods and services, inflation in wages and other costs, changes

in the exchange rate, changes in the unemployment rate, imported infla-
tion, etc. To avoid simultaneous relations, where a variable is affected
by inflation, all of these additional possible explanations will be entered
lagged. The choice of additional candidate variables is based on looser
theoretical guidelines, then some method of model selection is applied
to pick the 'best' relation between y_t and a subset of the $\{\mathbf{x}_t\}$ within a
class of functional connections (such as a linear relation with constant
parameters and small, identically-distributed errors e_t independent of the
$\{\mathbf{x}_t\}$). When N is very large ('big data', which could include micro-level
data on household characteristics or internet search data), most current
approaches have difficulties either in controlling the number of spurious
relationships that might be found (because of an actual or implicit signif-
icance level for hypothesis tests that is too loose for the magnitude of N),
or in retaining all of the relevant explanatory variables with a high prob-
ability (because the significance level is too stringent): see e.g., Doornik
and Hendry (2015). Moreover, the selected model may be hard to inter-
pret, and if many equations have been tried (but perhaps not reported),
the statistical properties of the resulting selected model are unclear: see
Leamer (1983).

6.2 The Drawbacks of Using Each Approach in Isolation

Many variants of theory-driven and data-driven approaches exist, often
combined with testing the properties of the e_t, the assumptions about
the regressors, and the constancy of the relationship $f(\cdot)$, but with dif-
ferent strategies for how to proceed if any of the conditions required for
viable inference are rejected. The assumption made all too often is that a
rejection occurs because that test has power under the specific alternative
for which the test is derived, although a given test can reject for many
other reasons. The classic example of such a 'recipe' is finding residual
autocorrelation and assuming it arose from error autocorrelation, whereas
the problem could be mis-specified dynamics, unmodelled location shifts

as seen above, or omitted autocorrelated variables. In our inflation example, in order to eliminate autocorrelation, annual dynamics need to be modelled, along with shifts due to wars, crises and legislative changes. The approach proposed in the next section instead seeks to include all likely determinants from the outset, and would revise the initial general formulation if any of the mis-specification tests thereof rejected.

Most observational data are affected by many influences, often outside the relevant subject's purview—as the 2016 Brexit vote has emphasized for economics—and it would require a brilliant theoretical analysis to take all the substantively important forces into account. Thus, a purely theory-driven approach, such as a monetary theory of aggregate inflation, is unlikely to deliver a complete, correct and immutable model that forges a new 'law' once estimated. Rather, to capture the complexities of real world data, features outside the theory remit almost always need to be taken into account, especially changes resulting from unpredicted events. Moreover, few theories include all the variables that characterize a process, with correct dynamic reactions, and the actual non-linear connections. In addition, the data may be mis-measured for the theory variables (revealed by revising national accounts data as new information accrues), and may even be incorrectly recorded relative to its own definition, leading to outliers. Finally, shifts in relationships are all too common—there is a distinct lack of empirical models that have stood the test of time or have an unblemished forecasting track record: see Hendry and Pretis (2016).

Many of the same problems affect a purely data-driven approach unless the $\mathbf{x}_t$ provide a remarkably comprehensive specification, in which case there will often be more candidate variables N than observations T: see Castle and Hendry (2014) for a discussion of that setting. Because included regressors will 'pick up' influences from any correlated missing variables, omitting important factors usually entails biased parameter estimates, badly behaved residuals, and most importantly, often non-constant models. Failing to retain relevant theory-based variables can be pernicious and potentially distort which models are selected. Thus, an approach that retains, but does not impose, theory-driven variables without affecting the estimates of a correct, complete, and constant theory model, has much to

offer, if it also allows selection over a much larger set of candidate variables, avoiding the substantial costs when relevant variables are omitted from the initial specification. We now describe how the benefits of the two approaches can be combined to achieve that outcome based on Hendry and Doornik (2014) and Hendry and Johansen (2015).

6.3   A Combined Approach

Let us assume that the theory correctly specifies the set of relevant variables. This could include lags of the variables to represent an equilibrium-correction mechanism. In the combined approach, the theory relation is retained while selecting over an additional set of potentially relevant candidate variables. These additional candidate variables could include disaggregates for household characteristics (in panel data), as well as the variables noted above. To ensure an encompassing explanation, the additional set of variables could also include additional lags and non-linear functions of the theory variables, other explanatory variables used by different investigators, and indicator variables to capture outliers and shifts.

The general unrestricted model (GUM) is formulated to nest both the theory model and the data-driven formulation. As the theory variables and additional variables are likely to be quite highly correlated, even if the theory model is exactly correct the model estimates are unlikely to be the same as those from estimating the theory model directly. However, the theory variables can be orthogonalized with respect to the additional variables, which means that they are uncorrelated with the other variables. Therefore, inclusion of additional regressors will not affect the estimates of the theory variables in the model, regardless of whether any, or all, of the additional variables are included. The theory variables are always included in the model, and any additional variables can be selected over to see if they are useful in explaining the phenomona of interest. Thus, data-based model selection can be applied to all the potentially relevant candidate explanatory variables while retaining the theory model without selection.

> **Summary of the Combined Approach**
> The theory variables are given by the set $\mathbf{z}_t$ of n relevant variables entering $f(\cdot)$. We use the explicit parametrization for $f(\cdot)$ of a linear, constant parameter vector $\boldsymbol{\beta}$, so the theory model is:
> $y_t = \boldsymbol{\beta}'\mathbf{z}_t + e_t$,
> where $e_t \sim \mathsf{IN}[0, \sigma_e^2]$ is independent of $\mathbf{z}_t$.
> Define the additional set of M candidate variables as $\{\mathbf{w}_t\}$.
>
> Formulate the GUM as:
> $y_t = \boldsymbol{\beta}'\mathbf{z}_t + \boldsymbol{\gamma}'\mathbf{w}_t + v_t$,
> which nests both the theory model and the data-driven formulation when $\mathbf{x}_t = (\mathbf{z}_t, \mathbf{w}_t)$, so v_t will inherit the properties of e_t when $\boldsymbol{\gamma} = \mathbf{0}$.
>
> Without loss of generality, $\mathbf{z}_t$ can be orthogonalized with respect to $\mathbf{w}_t$ by projecting the latter onto the former in:
> $\mathbf{w}_t = \boldsymbol{\Gamma}\mathbf{z}_t + \mathbf{u}_t$
> where $\mathsf{E}[\mathbf{z}_t\mathbf{u}_t'] = \mathbf{0}$ for estimated $\boldsymbol{\Gamma}$. Substitute the estimated components $\boldsymbol{\Gamma}\mathbf{z}_t$ and $\mathbf{u}_t$ for $\mathbf{w}_t$ in the GUM, leading to:
>
> $$y_t = \boldsymbol{\beta}'\mathbf{z}_t + \boldsymbol{\gamma}'(\boldsymbol{\Gamma}\mathbf{z}_t + \mathbf{u}_t) + v_t = (\boldsymbol{\beta}' + \boldsymbol{\gamma}'\boldsymbol{\Gamma})\mathbf{z}_t + \boldsymbol{\gamma}'\mathbf{u}_t + v_t.$$
>
> When $\boldsymbol{\gamma} = \mathbf{0}$, the coefficient of $\mathbf{z}_t$ remains $\boldsymbol{\beta}$, and because $\mathbf{z}_t$ and $\mathbf{u}_t$ are now orthogonal by construction, the estimate of $\boldsymbol{\beta}$ is unaffected by whether or not any or all $\mathbf{u}_t$ are included during selection.

To favour the incumbent theory, selection over additional variables can be undertaken at a stringent significance level to minimize the chances of spuriously selecting irrelevant variables. We suggest $\alpha = \min(0.001, 1/N)$. However, the approach protects against missing important explanatory variables, one such example of which is location shifts. The critical value for 0.1% in a Normal distribution is $c_{0.001} = 3.35$, so substantive regressors or shifts should still be easily retained. As noted in Castle et al. (2011), using IIS allows near Normality to be a reasonable approximation. However, a reduction from an integrated to a non-integrated representation requires non-Normal critical values,

another reason for using tight significance levels during model selection. In practice, unless the parameters of the theory model have strong grounds for being of special interest, the orthogonalization step is unnecessary since the same outcome will be found just by retaining the theory variables when selecting over the additional candidates. An example of retaining a 'permanent income hypothesis' based consumption function relating the log of aggregate consumers' expenditure, c, to logs of income, i, and lagged c, orthogonalized with respect to the variables in Davidson et al. (1978), denoted DHSY, is provided in Hendry (2018).

When should an investigator reject the theory specification? As there are M additional variables included in the combined approach (in addition to the n theory variables which are not selected over), on average αM will be significant by chance, so if $M = 100$ and $\alpha = 1\%$ (so $c_{0.01} = 2.6$), on average there will be one adventitiously significant selection. Thus, finding that one of the additional variables was 'significant' would not be surprising even when the theory model was correct and complete. Indeed, the probabilities that none, one and two of the additional variables are significant by chance are 0.37, 0.37 and 0.18, leaving a probability of 0.08 of more than two being retained. However, using $\alpha = 0.5\%$ ($c_{0.005} = 2.85$), these probabilities become 0.61, 0.30 and 0.08 with almost no probability of 3 or more being significant; and 0.90, 0.09 and <0.01 for $\alpha = 0.1\%$, in which case retaining 2 or more of the additional variables almost always implies an incomplete or incorrect theory model.

When the total number of theory variables and additional variables exceeds the number of observations in the data sample (so $M + n = N > T$), our approach can still be implemented by splitting the variables into feasible sub-blocks, estimating separate projections for each sub-block, and replacing these subsets by their residuals. The n theory variables are retained without selection at every stage, only selecting over the (putatively irrelevant) variables at a stringent significance level using a multi-path block search of the form implemented in the model selection algorithm *Autometrics* (see Doornik 2009; Doornik and Hendry 2018). When the initial theory model is incomplete or incorrect—a likely possibility for the inflation illustration here—but some of the additional variables are relevant to explaining the phenomenon of interest, then an improved empirical model should result.

6.4   Applying the Combined Approach to UK Inflation Data

Interpreting regression equations

The simplest model considered below relates two variables, the dependent variable y_t and the explanatory variable x_t, $t = 1, \dots, T$:

$$y_t = \beta_0 + \beta_1 x_t + u_t.$$

To conduct inference on this model, we assume that the innovations $u_1, \dots u_T$ are independent and Normally distributed with a zero mean and constant variance, $u_t \sim \mathsf{IN}\left[0, \sigma_u^2\right]$, and that the parameter space for the parameters of interest $(\beta_0, \beta_1, \sigma_u^2)$ is not restricted. These assumptions need to be checked for valid inference, which is done by tests for residual autocorrelation (F_{ar}), non-Normality (χ_{nd}^2), autoregressive conditional heteroskedasticity (ARCH: F_{arch}, see Engle 1982), heteroskedasticity (F_{Het}), and functional form (F_{RESET}). If the assumptions for valid inference are satisfied, then we can interpret β_1 as the effect of a one unit increase in x_t on y_t, or an elasticity if x and y are in logs.

We start from the simplest equation relating inflation (denoted $\Delta p_t = p_t - p_{t-1}$ so Δ signifies a difference) to broad money growth (i.e., Δm_t) where lower case letters denote logs, P is the UK price level and M is its broad money stock:

$$\Delta p_t = \beta_0 + \beta_1 \Delta m_t + e_t \qquad (6.1)$$

The two time series for annual UK data over 1874–2012 are shown in Fig. 6.1(a) and their scatter plot with a fitted regression line and the deviations therefrom in Panel (b).

At first sight, the hypothesis seems to have support: the two series are positively related (from Panel (b)) and tend to move together over time (from Panel (a)), although much less so after 1980. However, that leaves open the question of why: is inflation responding to money growth, or is more (less) money needed because the price level has risen (fallen)?
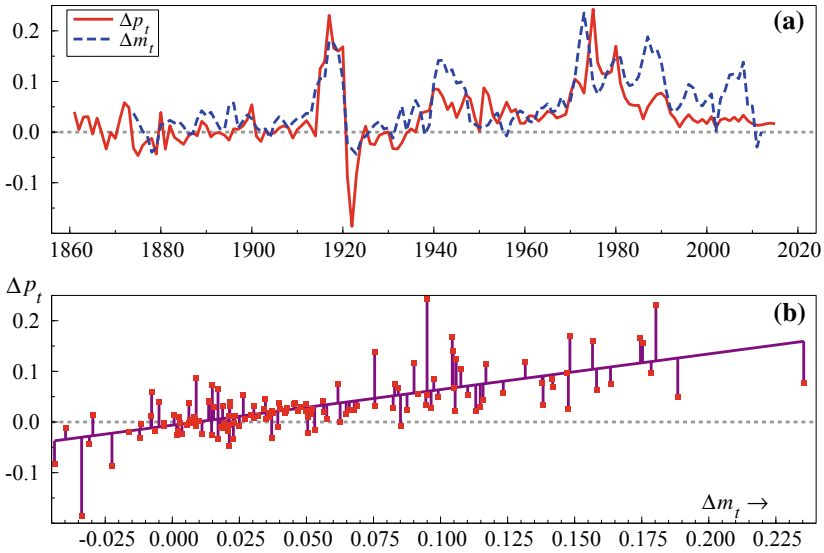
Fig. 6.1 (a) Time series of Δp_t and Δm_t; (b) scatter plot with the fitted regression of Δp_t on Δm_t and the deviations therefrom

The regression in (6.1) is estimated over 1877–2012 as:

$$\Delta p_t = \underset{(0.005)}{-\,0.005} + \underset{(0.063)}{0.69\,\Delta m_t} \tag{6.2}$$

$$\widehat{\sigma} = 4.1\% \quad \mathsf{R}^2 = 0.47 \quad \mathsf{F}_{ar}(2, 132) = 30.7^{**} \quad \mathsf{F}_{Het}(2, 133) = 4.35^*$$
$$\chi^2_{nd}(2) = 36.6^{**} \quad \mathsf{F}_{arch}(1, 134) = 8.94^{**} \quad \mathsf{F}_{RESET}(2, 132) = 1.27$$

The residual standard deviation, $\widehat{\sigma}$, is very large at 4%, with a 95% uncertainty range of 16%, when for the last 20 years, inflation has only varied between 1.5% and 3.5%.

Moreover, tests for residual autocorrelation (F_{ar}), non-Normality (χ^2_{nd}), autoregressive conditional heteroskedasticity (ARCH: F_{arch}) and heteroskedasticity (F_{Het}) all reject. Figure 6.2(a) records the fitted and actual values of Δp_t; (b) shows the residuals $\widehat{e}_t / \widehat{\sigma}$, (c) their density with a standard Normal for comparison; and (d) their residual correlogram.
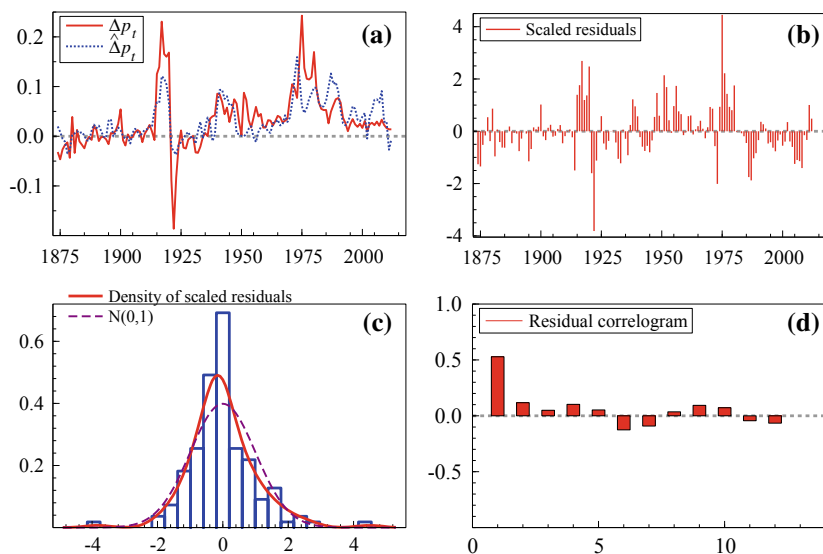
Fig. 6.2 (a) Fitted, $\widehat{\Delta p_t}$, and actual values, Δp_t, from (6.2); (b) scaled residuals $\widehat{e_t}/\widehat{\sigma}$; (c) their density with a standard Normal for comparison and (d) their residual correlogram

A glance at the test statistics in (6.2) and Fig. 6.2 shows that the equation is badly mis-specified, and indeed recursive estimation reveals considerable parameter non-constancy. The simplicity of the bivariate regression provides an opportunity to illustrate MIS, where both β_0 and β_1 are interacted with step indicators at every observation, so there are 271 candidate variables. Using $\alpha = 0.0001$ found 7 shifts in β_0 and 5 in β_1, halving $\widehat{\sigma}$ to 1.9%, and revealing a far from constant relationship between money growth and inflation.

Such a result should not come as a surprise given the large number of major regime changes impinging on the UK economy over the period as noted in Chapter 3, many relevant to the role of money. In particular, key financial innovations and changes in credit rationing included the introduction of personal cheques in the 1810s and the telegraph in the 1850s both reducing the need for multiple bank accounts just before our sample; credit cards in the 1950s; ATMs in the 1960s; deregulation of banks and building societies (the equivalent of US Savings and Loans) in the 1980s;

interest-bearing chequing accounts around 1984; and securitization of mortgages; etc.

First, to offer the incumbent theory a better chance, we added lagged values of Δp_{t-i} and Δm_{t-i} for $i = 1, 2$ to (6.2), but without indicators, which improves the fit to $\widehat{\sigma} = 3.3\%$ although three significant misspecification tests remain as (6.3) shows.

$$
\begin{aligned}
\Delta p_t = \quad &\underset{(0.09)}{0.67}\,\Delta p_{t-1} \ - \ \underset{(0.08)}{0.19}\,\Delta p_{t-2} \ + \ \underset{(0.10)}{0.40}\,\Delta m_t \\
&- \ \underset{(0.15)}{0.01}\,\Delta m_{t-1} \ - \ \underset{(0.11)}{0.02}\,\Delta m_{t-2} \ - \ \underset{(0.004)}{0.003}
\end{aligned}
\tag{6.3}
$$

$$
\widehat{\sigma} = 3.3\% \quad \mathsf{R}^2 = 0.66 \quad \mathsf{F}_{\mathsf{ar}}(1, 128) = 0.20 \quad \mathsf{F}_{\mathsf{Het}}(10, 125) = 5.93^{**}
$$

$$
\chi^2_{\mathsf{nd}}(2) = 76.9^{**} \quad \mathsf{F}_{\mathsf{arch}}(1, 134) = 7.73^{**} \quad \mathsf{F}_{\mathsf{RESET}}(2, 128) = 0.13
$$

Neither lag of money growth is relevant given the contemporaneous value, but both lags of inflation matter, suggesting about half of past inflation is carried forward, so there is a moderate level of persistence. Now applying IIS+SIS at $\alpha = 0.0025$ to (6.3) yielded $\widehat{\sigma} = 1.6\%$ with 4 impulse and 6 step indicators retained, but with all the coefficients of the economics variables being much closer to zero.

As the aim of this section is to illustrate our approach, and a substantive model of UK inflation over this period is available in Hendry (2015), we just consider four of the rival explanations that have been proposed. Thus to create a more general GUM for Δp_t, we also include the unemployment rate $(U_{r,t})$ relating to the original Phillips curve model of inflation (Phillips 1958); the potential output gap (measured by $(g_t - 0.019t)$ and adjusted to give a zero mean) and growth in GDP (Δg_t) to represent excess demand for goods and services (an even older idea dating back to Hume); wage inflation (Δw_t) as a cost push measure (a 1970s theme); and changes in long-term interest rates $(\Delta R_{L,t})$ reflecting the cost of capital. To avoid simultaneity, all variables are entered lagged one and two periods (including money growth) and the 2-period lag of the potential output gap is excluded to avoid multicollinearity between growth in GDP and the potential output gap, making $N = 14$ including the intercept before any indicators. The five additional variables are then orthogonalized with

respect to Δm_t and lags of it and lags of Δp_t. To fully implement the strategy, lags of regressors should also be orthogonalized, but the resulting coefficients of the variables in common are close to those in the simpler model. Estimation delivers $\hat{\sigma} = 3.3\%$ with an F-test on the additional variables of $F(9, 121) = 2.66^{**}$, thereby rejecting the simple model, still with the three mis-specification tests significant.

Since the baseline theory model is untenable, its coefficients are not of interest, so we revert to the original measures of all the economic variables to facilitate interpretation of the final model. The economic variables are all retained while we select indicators by IIS and SIS at $\alpha = 0.0025$, choosing that significance level so that only a few highly significant indicators would be retained, with almost none likely to be significant by chance (with a theoretical retention rate of $271 \times 0.0025 = 0.68$ of an indicator). Nevertheless, five impulse and ten step indicators were selected, producing $\hat{\sigma} = 1.2\%$ now with no mis-specification tests significant at 1%. Such a plethora of highly significant indicators implies that inflation is not being well explained even by the combination of all the theories. In fact the more general model in Hendry (2015) still needed 7 step indicators (though for somewhat different unexplained shifts) as well as dummies for the World Wars: sudden major shifts in UK inflation are not well explained by economic variables. We then selected over the 13 economic variables at the conventional significance level of 5%, forcing the intercept to be retained. Six were retained, with $\hat{\sigma} = 1.2\%$ to deliver (only reporting the economic variables):

$$\Delta p_t = \underset{(0.04)}{0.17}\,\Delta m_{t-2} - \underset{(0.10)}{0.46}\,U_{r,t-1} + \underset{(0.09)}{0.43}\,U_{r,t-2} - \underset{(0.03)}{0.10}\,gap_{t-1}$$

$$+ \underset{(0.03)}{0.23}\,\Delta w_{t-1} + \underset{(0.24)}{0.54}\,\Delta R_{L,t-1} + \underset{(0.01)}{0.02} \tag{6.4}$$

$$\hat{\sigma} = 1.2\% \quad (R^*)^2 = 0.96 \quad F_{ar}(2, 112) = 0.14$$
$$F_{Het}(22, 108) = 1.97^*$$
$$\chi^2_{nd}(2) = 1.63 \quad F_{arch}(1, 134) = 0.15$$
$$F_{RESET}(2, 112) = 4.25^*$$

In contrast to the simple monetary theory of inflation, the model retains aspects of all the theories posited above. Now there is no direct persistence from past inflation, but remember that the step indicators represent persistent location shifts, so the mean inflation rate persists at different

levels. Interesting aspects are how many shifts were found and that these location shifts seem to come from outside economics. The dates selected are consistent with that: 1914, 1920 and 1922, 1936 and 1948, 1950 and 1952, 1969, 1973 and 1980, all have plausible events, although they were not the only large unanticipated shocks over the last 150 years (e.g., the general strike). There is a much bigger impact from past wage growth than money growth as proximate determinants, but we have not modelled those to determine 'final' causes of what drives the shifts and evolution. Finally, the $-$ then $+$ coefficients on unemployment suggest it is changes therein rather than the levels that affect inflation.

The long-run relation after solving out the dynamics is:

$$\Delta p = \underset{(0.03)}{0.23\,\Delta w} + \underset{(0.04)}{0.17\,\Delta m} - \underset{(0.07)}{0.03\,U_r} - \underset{(0.03)}{0.10\,gap} + \underset{(0.24)}{0.54\,\Delta R_L}$$

(6.5)

The first two signs and magnitudes are easily interpreted as higher wage growth and faster money growth raise inflation. The negative unemployment coefficient is insignificant, consistent with its role probably being through changes. The hard to interpret output gap could be reflecting omitted variables and changes in the cost of capital raising inflation.

It is easy to think of other variables that could have an impact on the UK inflation rate, including the mark-up over costs used by companies to price their output; changes in commodity prices, especially oil; imported inflation from changes in world prices; changes in the nominal exchange rate; and changes in the National Debt among others, several of which are significant in the inflation model in Hendry (2015). Moreover, there is no strong reason to expect a constant relation between any of the putative explanatory variables and inflation given the numerous regime shifts that have occurred, the changing nature of money, and increasing globalization. In principle, MIS could be used where shifts are most likely, but in practice might be hard to implement at a reasonable significance level.

In our proposed combined theory-driven and data-driven approach, when the theory is complete it is almost costless in statistical terms to check the relevance of large numbers of other candidate variables, yet there is a good chance of discovering a better empirical model when the theory is incomplete or incorrect. Automatic model selection algorithms that allow

retention of theory variables while selecting over many orthogonalized candidate variables can therefore deliver high power for the most likely explanatory variables while controlling spurious significance at a low level. Oh for having had the current technology in the 1970s! This is only partly anachronistic, as the theory in Hendry and Johansen (2015) could easily have been formulated 50 years ago. Combining the theory and data based approaches improves the chances of discovering an empirically well-specified, theory-interpretable model.

References

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

Castle, J. L., and Hendry, D. F. (2014). Semi-automatic non-linear model selection. In Haldrup, N., Meitz, M., and Saikkonen, P. (eds.), *Essays in Nonlinear Time Series Econometrics*, pp. 163–197. Oxford: Oxford University Press.

Davidson, J. E. H., Hendry, D. F., Srba, F., and Yeo, J. S. (1978). Econometric modelling of the aggregate time-series relationship between consumers' expenditure and income in the United Kingdom. *Economic Journal*, **88**, 661–692.

Doornik, J. A. (2009). Autometrics. In Castle, J. L., and Shephard, N. (eds.), *The Methodology and Practice of Econometrics*, pp. 88–121. Oxford: Oxford University Press.

Doornik, J. A., and Hendry, D. F. (2015). Statistical model selection with big data. *Cogent Economics and Finance*. http://www.tandfonline.com/doi/full/10.1080/23322039.2015.1045216#.VYE5bUYsAsQ.

Doornik, J. A., and Hendry, D. F. (2018). *Empirical Econometric Modelling Using PcGive: Volume I*, 8th Edition. London: Timberlake Consultants Press.

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

Hendry, D. F. (2015). *Introductory Macro-econometrics: A New Approach*. London: Timberlake Consultants. http://www.timberlake.co.uk/macroeconometrics.html.

Hendry, D. F. (2018). Deciding between alternative approaches in macro-economics. *International Journal of Forecasting*, **34**, 119–135, with

'Response to the Discussants', 142–146. http://authors.elsevier.com/sd/article/S0169207017300997.

Hendry, D. F., and Doornik, J. A. (2014). *Empirical Model Discovery and Theory Evaluation.* Cambridge, MA: MIT Press.

Hendry, D. F., and Johansen, S. (2015). Model discovery and Trygve Haavelmo's legacy. *Econometric Theory*, **31**, 93–114.

Hendry, D. F., and Pretis, F. (2016). *All Change! The Implications of Non-stationarity for Empirical Modelling, Forecasting and Policy.* Oxford University: Oxford Martin School Policy Paper. http://www.oxfordmartin.ox.ac.uk/publications/view/2318.

Leamer, E. E. (1983). Let's take the con out of econometrics. *American Economic Review*, **73**, 31–43. Reprinted in Granger, C. W. J. (ed.) (1990), *Modelling Economic Series.* Oxford: Clarendon Press.

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. Reprinted as pp. 243–260 in R. Leeson (ed.) (2000) *A. W. H. Phillips: Collected Works in Contemporary Perspective.* Cambridge: Cambridge University Press.