

Nowcasting from Disaggregates in the Face of Location Shifts

Jennifer L. Castle and David F. Hendry*
Department of Economics, Oxford University.

September 9, 2009

Abstract

Given a need for nowcasting, we consider how nowcasts can best be achieved, the use and timing of information, including disaggregation over variables and common features, and the role of automatic model selection for nowcasting missing disaggregates. We focus on the impact of location shifts on nowcast failure and nowcasting during breaks, using impulse saturation, its relation to intercept correction, and to robust methods to avoid systematic nowcast failure. We propose a nowcasting strategy, building models of all N disaggregate series by automatic methods, forecasting every variable each period, then testing for shifts in available measures, switching to robust forecasts of missing series when breaks are detected.

Keywords: Nowcasting; disaggregate information; Autometrics; location shifts; impulse saturation; robust forecasts.

JEL Classification: C51; C52.

1 Introduction

Given the need for nowcasts, an issue discussed extensively in Clements and Hendry (2003) and Castle, Fawcett and Hendry (2009d), we first consider in section 2 how they might best be produced. Section 3 then discusses the aggregation of disaggregate information, building on Hendry and Hubrich (2009). Since large numbers of time series are involved, and random subsets are missing intermittently, section 4 discusses the potential role of automatic model selection for nowcasting all the disaggregates. Our approach seeks to handle all the available information allowing for multiple past breaks at unknown dates, and contemporaneous location shifts, so inherently involves more variables than observations in the proposed models. To do so, we extend the approach described in Castle, Doornik and Hendry (2009a), based on

*Corresponding author: Jennifer L. Castle, Department of Economics, Manor Road Building, Manor Road, Oxford, OX1 3UQ. Tel: +44 1865 271089, Fax: +44 1865 271094, E-mail: jennifer.castle@nuffield.ox.ac.uk.

Autometrics with impulse saturation: see Doornik (2009), Hendry, Johansen and Santos (2008), and Johansen and Nielsen (2009). Section 5 describes how these ideas can be combined for nowcasting the disaggregates. We then address how to handle location shifts in section 6: subsections 6.1 and 6.2 respectively discuss robust nowcasting and nowcasting during a break, leading to the description of the proposed nowcasting strategy in section 7. Section 8 concludes.

2 Producing ‘good’ nowcasts

A timely data source that provided an accurate ‘measure’ of a required aggregate variable is clearly preferable to seeking ‘good forecasts’ of it. Nevertheless, as there will usually be a role for forecasting ‘preliminary estimates’, the best use of forecasting methods is important, albeit that it is an empirical issue as to which methods might be ‘best’. Nowcasting is partly a ‘signal extraction’ problem for missing data entering the aggregate, but if the first announcements can be systematically improved by forecasting them either directly or via the disaggregates, then the quality of the resulting data are bound to be better. Four fundamental problems inhibit achieving that outcome.

First, the objective functions of the users of a nowcaster’s output are almost always unknown. A convenient approximation is to assume a quadratic loss, in which case the aim of nowcasting becomes to find a forecast $\hat{y}_{T|T-\delta}$ of the aggregate y_T which solves:

$$\underset{\hat{y}_{T|T-\delta}}{\operatorname{argmin}} E_T [y_T - \hat{y}_{T|T-\delta}]^2 \quad (1)$$

where $\hat{y}_{T|T-\delta} = g_{T-\delta}(\hat{\mathcal{J}}_{T-\delta})$, and $g_{T-\delta}(\cdot)$ is the relevant function of the available measured information set, $\hat{\mathcal{J}}_{T-\delta}$, which estimates the actual information, $\mathcal{I}_{T-\delta}$, on which y_T depends via $y_T = f_T(\mathcal{I}_{T-\delta})$. We have dated information as $T - \delta$ for $\delta > 0$, since some evidence must be unavailable at T to necessitate nowcasting. To proceed, we will assume that (1) does indeed provide the objective function, as it is difficult to see why either direction of asymmetry should dominate.

The second difficulty is that the information to be included in $\hat{\mathcal{J}}_{T-\delta}$ is also unknown, and could comprise the history of the series alone (via a univariate time-series model), survey information, current and past data on other related time series, past information about revisions etc. We will address this issue below by proposing automatic model selection either directly for the aggregate itself, or for nowcasting the disaggregates allowing for all the available information, multiple past breaks and contemporaneous location shifts as in Castle *et al.* (2009a), then relating the aggregate to its own past and all the disaggregates as in Hendry and Hubrich (2009). Section 3 considers relating aggregates to disaggregates, and §4 discusses automatic model selection.

The third, and most serious, problem is obtaining $\hat{y}_{T|T-\delta}$. At first sight, it may be thought that the conditional expectation $E_T[y_T|\hat{\mathcal{J}}_{T-\delta}]$ of y_T given $\hat{\mathcal{J}}_{T-\delta}$ should be the solution to (1), namely the minimum mean-square error predictor. Unfortunately, $E_T[\cdot]$ is not known at T either, since the statistical process generating $\{y_t\}$ is also never known, and in economics is always wide-sense non-stationary, namely its distribution changes over time from both stochastic trends

and location shifts, reflected in our notation $f_T(\cdot)$. Thus, the form of the optimal predictor $\hat{y}_{T|T-\delta}$ is never known. Worse, $E_{T-\delta}[y_T|\hat{\mathcal{J}}_{T-\delta}]$, which might be available, need not be a good forecast device when location shifts occur, as it is calculated over the wrong distribution: see e.g., Castle, Doornik, Hendry and Nymoen (2009b). Thus, the realized mean square forecast error (MSFE) given (1) is:

$$E_T \left[f_T(\mathcal{I}_{T-\delta}) - g_{T-\delta}(\hat{\mathcal{J}}_{T-\delta}) \right]^2 \quad (2)$$

leading to the following abbreviated taxonomy of forecast errors $\hat{u}_{T|T-\delta} = y_T - \hat{y}_{T|T-\delta}$:

$$\begin{aligned} \hat{u}_{T|T-\delta} &= f_T(\mathcal{I}_{T-\delta}) - g_{T-\delta}(\hat{\mathcal{J}}_{T-\delta}) \\ &= f_T(\mathcal{I}_{T-\delta}) - f_{T-\delta}(\mathcal{I}_{T-\delta}) - \text{distribution shift (§6)} \\ &\quad + f_{T-\delta}(\mathcal{I}_{T-\delta}) - g_{T-\delta}(\mathcal{I}_{T-\delta}) - \text{model mis-specification (§4)} \\ &\quad + g_{T-\delta}(\mathcal{I}_{T-\delta}) - g_{T-\delta}(\mathcal{J}_{T-\delta}) - \text{reduced information (§3.1)} \\ &\quad + g_{T-\delta}(\mathcal{J}_{T-\delta}) - g_{T-\delta}(\hat{\mathcal{J}}_{T-\delta}) - \text{measurement error} \end{aligned} \quad (3)$$

Clements and Hendry (1998, 1999) provide a framework for analyzing the properties of forecasting models in wide-sense non-stationary processes, when the device being used does not coincide with the generating mechanism of the process (Clements and Hendry, 2008, provide a non-technical explanation). To summarize, they explain:

- (a) the recurrent episodes of systematic mis-forecasting that have occurred historically;
 - (b) show that *ex ante* well-specified models need not forecast better than badly specified; since
 - (c) causally-relevant variables need not improve forecasting over irrelevant variables;
 - (d) the benefits of many of the empirical practices of forecasters, such as intercept corrections;
 - (e) show why pooling across a range of methods and models can be beneficial (see e.g., Hendry and Clements, 2004), but need not be unless carefully undertaken (see Hendry and Reade, 2008);
 - (f) why forecasting devices that are robust to location shifts do not experience systematic forecast failure; and so
 - (g) dominate in forecasting competitions (such as Makridakis and Hibon, 2000: see Fildes and Ord, 2002, and Clements and Hendry, 2001); hence
 - (h) why so-called ‘naive devices’ (simple adaptive devices like damped trend, differenced models, and exponentially weighted moving averages–EWMA) can outperform; whereas
 - (i) equilibrium-correction models can experience systematic forecast failure (see Hendry, 2006).
- Despite its non-specific assumptions, therefore, a theory of forecasting which allows for unanticipated structural breaks in an evolving economic mechanism for which the forecasting model is mis-specified in unknown ways can provide a useful basis for interpreting, and potentially circumventing, systematic forecast failure in economics. These implications seem to carry over to nowcasting, so section 6 considers the issue of breaks in that context.

Fourthly, the disaggregates that have been measured at T as partial inputs to calculating y_T are themselves subject to unknown inaccuracies and future revisions. This is more difficult to handle, but Castle *et al.* (2009d) seek to address the impact of measurement errors in the

context of the much simpler EWMA model, which applies to a random walk that is subject to measurement errors.

All these considerations seem to apply to nowcasting an aggregate either directly or via its disaggregates.

3 Aggregation

The variable of interest, y_T , say GDP, is an aggregate variable, comprising $y_t = \sum_{i=1}^N w_i y_{i,t}$ where $y_{i,t}$ are the disaggregates and w_i are the weights, which could be changing over time. Data are released with varying time delays, such that at time T some components of y_T will be observed and some will be unavailable until $T + \delta$. For $y_{i,T}$, the $i = 1, \dots, J$, components are known at T and the $i = J + 1, \dots, N$, components are unknown at T . The two sets are not uniquely defined, so an individual element can switch back and forth between the sets through time depending on how the information is accrued and J is not fixed. Hence, a forecasting strategy needs to be flexible enough to allow for changes in the timing of releases. As some components of y_T are unknown at T a nowcast is computed. There are three alternative methods, see Hendry and Hubrich (2009):

- I Forecast the aggregate using only aggregate information, some of which perforce must be lagged one period:

$$\hat{y}_{T|T-\delta^*} = f(\mathbf{Y}_{T-1}^0, \mathbf{Z}_{T-\delta^*}^0),$$

where $\mathbf{Y}_{T-1}^0 = y_{T-1}, \dots, y_0$, and $\mathbf{Z}_{T-\delta^*}^0 = \mathbf{z}_{T-\delta^*}, \dots, \mathbf{z}_0$ is a vector of conditioning variables such as surveys or leading indicators, which may be more recent than the latest aggregate observation, with $\delta^* \leq \delta \leq 1$.

- II Forecast those disaggregates that are unknown at T :

$$\tilde{y}_{i,T|T-\delta^*} = f(\mathbf{y}_{i,T-\delta}^0, \mathbf{Z}_{T-\delta^*}^0), \quad i = J + 1, \dots, N,$$

then aggregate their forecasts together with the known data:

$$\tilde{y}_{T|T} = \sum_{i=1}^J w_i y_{i,T} + \sum_{i=J+1}^N w_i \tilde{y}_{i,T|T-\delta^*}$$

Typically the weights are taken as given but could also be forecast; we abstract from the issue of weights in the subsequent analysis, as with GDP data.

- III Forecast the aggregate, conditioning on both aggregate and disaggregate information. There are two possible methods.

- a. In order to apply standard forecasting procedures, a balanced panel is needed:

$$\bar{y}_{T|T-\delta} = f(\mathbf{Y}_{T-\delta}^0, \mathbf{y}_{i,T-\delta}^0, \mathbf{Z}_{T-\delta}^0), \quad \forall i \in N,$$

where the conditioning information includes information dated $T - \delta$ and previous, resulting in a δ -step ahead forecast.

- b. Alternatively, an unbalanced panel can be used, conditioning on whatever information is also available at T :

$$\bar{\bar{y}}_{T|T} = f \left(\mathbf{Y}_{T-1}^0; \mathbf{y}_{1,T}^0, \dots, \mathbf{y}_{J,T}^0, \mathbf{y}_{J+1,T-\delta}^0, \dots, \mathbf{y}_{N,T-\delta}^0; \mathbf{Z}_{T-\delta}^0 \right),$$

resulting in a large unbalanced panel in which there are missing observations at the end of the sample: see Wallis (1986) for a discussion of the ‘ragged edge’ problem.

3.1 Nowcasting aggregates from disaggregates

[I] has been the most common approach in the literature, with a focus on predicting the revision process at the aggregate level, see Lee, Olekalns and Shields (2008). However, the aggregate approach relies on a reduced information set, $\mathcal{J}_{T-1} \subset \mathcal{I}_{T-1}$, where \mathcal{I}_{T-1} is the full information set including disaggregate information. As unpredictability is relative to the information set used, using a subset of information will result in less accurate, albeit unbiased, predictions, see Clements and Hendry (1998, p.36). Hence, we do not consider [I] in this analysis, instead focusing on using the additional disaggregate information. When all disaggregates are observed, Mayo and Espasa (2009) discuss how to incorporate cointegrated and cyclical information about disaggregates to improve the aggregate forecasts relative to unrestricted estimation.

We now show analytically that when interest focuses on predicting the aggregate, then nothing is lost by doing so directly from disaggregate information, without predicting the disaggregates, i.e., method [III]. $N = 2$ suffices to illustrate the analysis, which generalizes to many components:

$$y_T = w_1 y_{1,T} + (1 - w_1) y_{2,T} \quad (4)$$

with weights w_1 and $w_2 = (1 - w_1)$, where for simplicity we take the weights to be constant over time, as in a simple sum aggregate. Assume that $J = 0$, so the contemporaneous disaggregates are not observed, and information is only available from $T - 1$. The DGP for the disaggregates is assumed to be:

$$y_{i,t} = \gamma_i' \mathbf{x}_{t-1} + \eta_{i,t}$$

where $\mathbf{x}_{t-1} = (y_{1,t-1}, \dots, y_{N,t-1}; \mathbf{z}_{t-1})'$ denotes all the available information, which includes the lagged disaggregates, so:

$$\mathbb{E}_{T-1} [y_{i,T} | \mathbf{x}_{T-1}] = \gamma_i' \mathbf{x}_{T-1} \quad \text{for } i = 1, 2 \quad (5)$$

Aggregating the two terms in (5), delivers:

$$\mathbb{E}_{T-1} [y_T | \mathbf{x}_{T-1}] = \sum_{i=1}^2 w_i \mathbb{E}_{T-1} [y_{i,T} | \mathbf{x}_{T-1}] = \left(\sum_{i=1}^2 w_i \gamma_i' \right) \mathbf{x}_{T-1} = \boldsymbol{\psi}' \mathbf{x}_{T-1} \quad (6)$$

say. Predicting y_T directly from \mathbf{x}_{T-1} yields:

$$\mathbb{E}_{T-1} [y_T | \mathbf{x}_{T-1}] = \boldsymbol{\pi}' \mathbf{x}_{T-1} \quad (7)$$

Since the left-hand sides of (6) and (7) are equal:

$$\pi' \mathbf{x}_{T-1} = \psi' \mathbf{x}_{T-1} \quad (8)$$

so nothing is lost by predicting y_T directly, instead of aggregating component predictions once the same information set \mathbf{x}_{T-1} is used for both, i.e. there is no benefit to method [II] over [IIIa].

Hendry and Hubrich (2009) show that this analysis generalizes to non-constant weights and changing parameters, and demonstrate that forecast-origin breaks equally affect combining disaggregate forecasts to forecast an aggregate as in (6) or forecasting the aggregate directly as in (7), so there is no benefit in MSFE terms from building models of disaggregates. However, that analysis assumes that all conditioning information is available at $T - 1$, and does not make use of the contemporaneous information available at T that is an essential component of the nowcasting strategy. Discarding information available at T is particularly costly when breaks occur simultaneously across the disaggregates (or a subset) as this information can be used to ensure robust forecasts in the presence of breaks. This implies that either [II] or [IIIb] is a preferable strategy.

There is a range of methods for unbalanced panels in which all available data are used, including dynamic factor models, see Giannone, Reichlin and Small (2008) and Schumacher and Breitung (2008), and mixed frequency time series models such as MIDAS, see Kuzin, Marcellino and Schumacher (2009). Marcellino and Schumacher (2007) combine factor models with MIDAS, and Ferrara, Guegan and Rakotomalahy (2008) use non-parametric methods, based on nearest neighbors and on radial basis function approaches to nowcast unbalanced monthly data sets. However, these methods do not take account of structural breaks in the disaggregates, instead abstracting from wide-sense nonstationarity. Hence, the approach we propose is to use [II], but augmented for structural break detection, an issue to which we now turn.

4 Automatic model selection with more variables than observations

Automatic model selection for nowcasting all disaggregates allowing for all the available information, multiple past breaks and contemporaneous location shifts can be undertaken following the approach described in Castle *et al.* (2009a), based on *Autometrics* with impulse saturation (see Doornik, 2009, on the former, and Hendry *et al.*, 2008, and Johansen and Nielsen, 2009, for the latter). The theory of reduction (see e.g. Hendry, 2009, for a recent exposition) explains the existence of a local data generation process (LDGP) for any subset of variables, and the relationship of any model to that LDGP. The idea behind general-to-specific (Gets) model selection is to locate a good approximation to that LDGP, characterized by its being a congruent representation and encompassing the evidence on the LDGP directly and via that embodied in other models. To do so, one commences from the most general model that nests all the candidate variables, their lags, functional form transformations and possible breaks, intrinsically leading to having more variables, N , than observations T . *Autometrics* handles the $N > T$ problem

by a mixture of expanding and contracting searches that seek all the variables relevant at the chosen significance level α , set such that αN remains small (e.g., unity). Multiple breaks are tackled conjointly by impulse saturation (IS) which adds an impulse indicator for every observation: Hendry *et al.* (2008) establish one feasible algorithm, and derive the null distribution for an IID process, and Johansen and Nielsen (2009) generalize their findings to general dynamic regression models (possibly with unit roots), and show that there is a very small efficiency loss under the null of no breaks when αT is small, despite investigating the potential relevance of T additional variables. Castle *et al.* (2009a) examine the ability of IS to detect multiple breaks, and show it can find up to 20 breaks in 100 observations.

The intent of Gets is to find ‘good’ LDGP models as defined above, but as noted in (b) in section 2, Clements and Hendry (1998, 1999) show that one cannot prove that the pre-existing LDGP is the best model for forecasting. We address how to handle that issue in section 6 below.

5 Nowcasting the disaggregates

Nowcasts are required of the $N - J$ unknown disaggregates. As breaks may have occurred in the individual series over the sample period, it is important to account for these in the nowcasting models. The in-sample general unrestricted model (GUM) for nowcasting the unknown disaggregates is given by:

$$y_{i,t} = \beta' \mathbf{x}_{t-1} + \alpha' \mathbf{f}_{t-1} + \sum_{j=1}^{T-1} \varsigma_j 1_{t_j=t} + \nu_t, \quad (9)$$

for $t = 1, \dots, T-1$, and $i = J+1, \dots, N$. \mathbf{x}_{t-1} denotes all the available information including the lagged disaggregates and additional explanatory variables such as survey information or leading indicators, \mathbf{f}_{t-1} denotes a set of q latent common factors (discussed in section 5.1), and $1_{t_j=t}$ is a set of $T-1$ individual impulses taking the value unity at t and zero otherwise. Selection is undertaken using *Autometrics* and the resulting forecasts are given by:

$$\tilde{y}_{i,T|T-1} = \hat{\gamma}' \mathbf{x}_{T-1}^* + \hat{\theta}' \mathbf{f}_{T-1}^* + \hat{\zeta}' \mathbf{d}_{T-1} \quad (10)$$

where \mathbf{x}_{T-1}^* and \mathbf{f}_{T-1}^* denote the retained variables and factors after selection and \mathbf{d}_{T-1} collects the retained impulses from selection of (9). Loose significance levels could be used to select the forecasting model regressors in \mathbf{x}_{T-1}^* : Clements and Hendry (2005) show that all regressors with squared non-centralities greater than unity should be retained. However, tight significance levels should be used to select the impulses as $\alpha (T-1)$ impulses will be retained on average under the null from (9). In practice, *Autometrics* imposes one significance level so selection could be undertaken using a 2-step procedure by appealing to the Frisch and Waugh (1933) theorem, first identifying the impulses and then selecting the regressors conditional on the impulses.

5.1 Common features

The disaggregate series are likely to have common components, such as cycles and trends, as well as idiosyncratic elements, so it can be useful to include these in the GUM: see Mayo and Espasa (2009). As N is large, it may be difficult to identify (say) cointegrating relationships across all disaggregates, but factors can proxy missing cointegration information. There are many ways to obtain the factors f_i for $i = 1, \dots, q$: Stock and Watson (2002) suggest using static principal component analysis applied to $\mathbf{y}_d = (\mathbf{y}_1, \dots, \mathbf{y}_N)'$:

$$\mathbf{f}_t = \hat{\mathbf{\Lambda}}' \mathbf{y}_{d,t}$$

where the set of q factor loadings are collected in $\hat{\mathbf{\Lambda}}$, the $N \times q$ matrix of eigenvectors corresponding to the q largest eigenvalues of the sample covariance matrix $\hat{\mathbf{\Sigma}}$. There are alternative methods of estimating the factors: e.g., Forni, Hallin, Lippi and Reichlin (2005) propose a weighted version of the static principal components estimator, where time series are weighted according to their signal-to-noise ratio. However, static principal components capture the main aspects of common trends and cycles, and seem sufficient for the purpose of nowcasting currently missing disaggregates. As the factors mainly capture long-run relationships, the available information on the J known disaggregates at T will not contribute much, so the factors can be calculated over the full sample of N to $T - 1$.

In practice, we propose separating the disaggregates into blocks corresponding to groups of variables that exhibit common trends or cycles, as in section 7. The factors would be obtained for each block, f_i^j , where $i = 1, \dots, q$ denotes the factor and $j = 1, \dots, b$ denotes the block $\mathbf{y}_{d,j}$ where \mathbf{y}_d is divided into b subsets, but all factors can be included in the GUM. The key innovation is that both individual explanatory variables and the factors are included jointly in the GUM, in contrast to much of the dynamic factor models literature. This is feasible due to the ability of *Autometrics* to handle more variables than observations, and hence perfect collinearity. If factors are helpful in explaining movements in the disaggregates, they should be retained in the selected model (10).

5.2 Using all available information

The problem with the forecasts from (10) is that additional information may be available through the disaggregates known at T , $y_{i,T}$ for $i = 1, \dots, J$, allowing for more rapid identification of outliers or location shifts. Hence, we propose an augmented forecast of (10), $\tilde{y}_{i,T|T}$. To motivate the nowcast, we first outline robust forecasting before proposing the nowcasting strategy.

6 Robust forecasting

To explain the principles, we draw on Hendry (2006). Consider forecasting from a first-order cointegrated vector equilibrium correction model (VEqCM) in n $I(1)$ variables \mathbf{x}_t given in-sample by:

$$\Delta \mathbf{x}_t = \gamma + \alpha (\beta' \mathbf{x}_{t-1} - \mu) + \epsilon_t \text{ where } \epsilon_t \sim \text{IN}_n [0, \Omega_\epsilon] \quad (11)$$

so $\Delta \mathbf{x}_t$ and $\beta' \mathbf{x}_t$ are $I(0)$, with $E[\beta' \mathbf{x}_t] = \boldsymbol{\mu}$, $E[\Delta \mathbf{x}_t] = \boldsymbol{\gamma}$, β is $r \times 1$, $\boldsymbol{\gamma}$ is $n \times 1$ (but subject to the r restrictions $\beta' \boldsymbol{\gamma} = 0$), and $\boldsymbol{\mu}$ is $r \times 1$, leaving n unrestricted intercepts in total in (11). When the parameters of (11) are constant in-sample and known, and all variables correctly measured, 1-step ahead forecasts coincide with the conditional expectation $E_T[\Delta \mathbf{x}_{T+1} | \mathbf{x}_T]$:

$$\Delta \hat{\mathbf{x}}_{T+1|T} = \boldsymbol{\gamma} + \boldsymbol{\alpha} (\beta' \mathbf{x}_T - \boldsymbol{\mu}) \quad (12)$$

with forecast error $\hat{\boldsymbol{\epsilon}}_{T+1|T} = \Delta \mathbf{x}_{T+1} - \Delta \hat{\mathbf{x}}_{T+1|T} = \boldsymbol{\epsilon}_{T+1}$. Uncertain forecast origin values $\hat{\mathbf{x}}_T$ and parameter estimation lead to:

$$\Delta \hat{\mathbf{x}}_{T+1|T} = \hat{\boldsymbol{\gamma}} + \hat{\boldsymbol{\alpha}} (\hat{\beta}' \hat{\mathbf{x}}_T - \hat{\boldsymbol{\mu}}) \quad (13)$$

where the latter add terms of $O_p(T^{-1})$, which are negligible relative to the impact of location shifts in the DGP, the issue we now consider.

6.1 Robust nowcasting

We first consider 1-step ahead forecasting, so let $\Delta \mathbf{x}_T = (y_{1,T}, \dots, y_{J,T})'$ be the vector of known disaggregates and assume a DGP given by (11). Although location shifts are the key problem, we allow all parameters of the DGP to shift at $T - 1$ and (11) to be mis-specified by omission of a set of relevant variables $\{\mathbf{z}_t\}$, so that:

$$\Delta \mathbf{x}_T = \boldsymbol{\gamma}_0^* + \boldsymbol{\alpha}_0^* ((\beta_0^*)' \mathbf{x}_{T-1} - \boldsymbol{\mu}_0^*) + \boldsymbol{\Upsilon}_0^* \mathbf{z}_T + \boldsymbol{\nu}_T \quad (14)$$

and from then on:

$$\Delta \mathbf{x}_{T+1} = \boldsymbol{\gamma}_0^* + \boldsymbol{\alpha}_0^* ((\beta_0^*)' \mathbf{x}_T - \boldsymbol{\mu}_0^*) + \boldsymbol{\Upsilon}_0^* \mathbf{z}_{T+1} + \boldsymbol{\nu}_{T+1} \quad (15)$$

The forecast errors from (13) are:

$$\hat{\boldsymbol{\nu}}_{T+1|T} = \boldsymbol{\gamma}_0^* - \hat{\boldsymbol{\gamma}} + \boldsymbol{\alpha}_0^* ((\beta_0^*)' \mathbf{x}_T - \boldsymbol{\mu}_0^*) - \hat{\boldsymbol{\alpha}} (\hat{\beta}' \hat{\mathbf{x}}_T - \hat{\boldsymbol{\mu}}) + \boldsymbol{\Upsilon}_0^* \mathbf{z}_{T+1} + \boldsymbol{\nu}_{T+1} \quad (16)$$

Every possible source of forecast error occurs in (16): stochastic and deterministic breaks (the shift to $*$ parameters), omitted variables (\mathbf{z}_{T+1}), inconsistent parameter estimates (given the other mis-specifications), estimation uncertainty (the $\hat{\cdot}$ on parameters in (13)), mis-measurement at the forecast origin ($\mathbf{x}_T - \hat{\mathbf{x}}_T$), and innovation errors ($\boldsymbol{\nu}_{T+1}$).

Surprisingly, (14) is simultaneously almost the solution to the very problem it creates. If one reads what that equation shows, then $\Delta \mathbf{x}_T$ contains nearly everything one ever wanted to know for forecasting, but was afraid to ask, namely $\boldsymbol{\gamma}_0^*$, $\boldsymbol{\alpha}_0^*$, β_0^* , $\boldsymbol{\mu}_0^*$, $\boldsymbol{\Upsilon}_0^*$ and even the unknown omitted variable \mathbf{z}_T . When forecasting, individual parameter estimates are merely instrumental in arriving at $\Delta \hat{\mathbf{x}}_{T+1|T}$, and (14) contains all the shifted parameter values as well as the unknown omitted variables. Thus, consider forecasting $\Delta \mathbf{x}_{T+1}$ simply by:

$$\Delta \tilde{\mathbf{x}}_{T+1|T} = \Delta \hat{\mathbf{x}}_T \quad (17)$$

then the forecast error $\tilde{\nu}_{T+1|T}$ is (assuming that after one-period revisions, $\hat{\mathbf{x}}_{T-1} = \mathbf{x}_{T-1}$):

$$\begin{aligned}\tilde{\nu}_{T+1|T} &= (\gamma_0^* + \alpha_0^* ((\beta_0^*)' \mathbf{x}_T - \mu_0^*) + \Upsilon_0^* \mathbf{z}_{T+1} + \nu_{T+1}) \\ &\quad - (\gamma_0^* + \alpha_0^* ((\beta_0^*)' \mathbf{x}_{T-1} - \mu_0^*) + \Upsilon_0^* \mathbf{z}_T + \nu_T) \\ &= \alpha_0^* (\beta_0^*)' \Delta \mathbf{x}_T + \Upsilon_0^* \Delta \mathbf{z}_{T+1} + \Delta \nu_{T+1}\end{aligned}\quad (18)$$

All terms in the last line are $I(-1)$, so will be ‘noisy’, but systematic forecast failure should not result. Differencing ν_{T+1} doubles the innovation error variance, whereas differencing \mathbf{z}_{T+1} could well reduce the overall error. Moreover, although $\alpha_0^* (\beta_0^*)' \Delta \mathbf{x}_T$ is unknown, $\hat{\alpha} \hat{\beta}' \Delta \hat{\mathbf{x}}_T$ is known and will usually be a better estimate than zero, leading to the forecasting device:

$$\Delta \bar{\mathbf{x}}_{T+1|T} = \Delta \hat{\mathbf{x}}_T + \hat{\alpha} \hat{\beta}' \Delta \hat{\mathbf{x}}_T \quad (19)$$

or:

$$\Delta^2 \bar{\mathbf{x}}_{T+1|T} = \Delta \left(\hat{\gamma} + \hat{\alpha} \left(\hat{\beta}' \hat{\mathbf{x}}_T - \hat{\mu} \right) \right)$$

which is the differenced VEqCM from (13).

Now consider $\Delta \mathbf{x}_T^\dagger = (y_{J+1,T}, \dots, y_{N,T})'$, the vector of unknown disaggregates. At T , the nowcast given by (17):

$$\Delta \tilde{\mathbf{x}}_{T|T-1}^\dagger = \Delta \hat{\mathbf{x}}_{T-1}^\dagger \quad (20)$$

will result in the nowcast error:

$$\begin{aligned}\tilde{\nu}_{T|T-1}^\dagger &= \left(\gamma_0^* + \alpha_0^* \left((\beta_0^*)' \mathbf{x}_{T-1}^\dagger - \mu_0^* \right) + \Upsilon_0^* \mathbf{z}_T + \nu_T \right) \\ &\quad - \left(\gamma + \alpha \left((\beta)' \mathbf{x}_{T-2}^\dagger - \mu \right) + \nu_{T-1} \right)\end{aligned}\quad (21)$$

which is the same forecast error as (16), abstracting from parameter estimation uncertainty and forecast origin uncertainty, apart from an additional error term, ν_{T-1} . Thus, at the break point, a forecasting model will fail to reflect the break—unless it is separately predicted—but robust forecasting models will rapidly adapt after a permanent break, and the forecasts will be unbiased at $T + 1$ as in (18). This suggests that robust forecasting devices could be useful in nowcasting when either a break is suspected (see e.g., Hendry, 2004), or is already detected by forecast-origin impulses. We next consider whether contemporaneous information available from the disaggregates can help to nowcast at a break origin.

6.2 Nowcasting during a break

A key problem is the rapid identification of outliers and/or location shifts in real time, to avoid distortionary effects on nowcasts: anticipation of imminent unusual events would be even more helpful in avoiding large errors, but is difficult to analyze. However, we address the former issue briefly, building on Castle, Fawcett and Hendry (2009c). Those authors consider the impacts of both external shifts, where the parameters of the model in use are unaltered, but data properties change—as in a change in collinearity—and internal shifts where the model itself alters. The former induces an immediate and unavoidable increase in the MSFE, but as the forecast origin

moves forward (e.g., to $T + 1 - \delta$), updating quickly mitigates that effect. Unfortunately, updating is not necessarily helpful when an internal break occurs, such as a change in the intercept. Moreover, even when the form and timing of the break are known to the modeller for a scalar process (e.g., an ogive starting at T), but the parameters governing the speed and magnitude are not, then the differenced VEqCM performs as well as the estimated break process, both in theory and practice.

Nevertheless, that result relies on the break being relatively persistent, such that the changes in parameters embedded in (15) reflect those current in the following period. For temporary shifts, the same strategy will exacerbate the forecast errors, so judgement is required as to which forecasting method to use after any form of shift, an issue that is especially pertinent when a ‘shift’ is merely a measurement error, an issue considered in Castle *et al.* (2009d).

6.3 Detecting breaks in the disaggregates

If a break occurs in more than one of the observed disaggregate series, information about the break will be contained in $y_{i,T}$, $i = 1, \dots, J$, which can be used to improve the forecast $\tilde{y}_{i,T|T-1}$, $i = J + 1, \dots, N$, by conditioning on contemporaneous information. The first stage is to detect whether a break has occurred in the disaggregates at time T . Conventional break detection tests such as Andrews (1993) and Bai and Perron (1998) will have low or no power as the break is at the end-point of the series, but impulse saturation can be used to detect structural breaks at T . A model analogous to (9) can be used, but with data available up to T :

$$y_{i,T} = \beta' \mathbf{x}_{t-1} + \rho' \mathbf{f}_{t-1} + \sum_{j=1}^T \varsigma_j 1_{t_j=t} + v_t, \quad (22)$$

for $i = 1, \dots, J$, and $t = 1, \dots, T$. If there is evidence of a break at T , $\hat{\varsigma}_T$ will be significant and the impulse will be retained. We suggest using a significance level of $0.5\% \leq \alpha \leq 0.1\%$ to detect the contemporaneous impulses. Such a conservative strategy will ensure that robust forecasts will only be used if there is strong evidence of a break in the nowcast period. The forecast error for the known disaggregates will also be informative:

$$e_{i,T} = y_{i,T} - \tilde{y}_{i,T|T-1}, \quad i = 1, \dots, J,$$

where $\tilde{y}_{i,T|T-1}$ would be obtained from the forecasting model (10). If $e_{i,T}$ is large relative to past $e_{i,t}$, there is evidence of a break in that disaggregate model. Whether breaks in the known disaggregate series help to predict breaks in the unknown disaggregate series is a question we address in section 7.

7 Nowcasting strategy

We can now summarize the ingredients of our proposed nowcasting strategy. 1-step ahead forecasts, $\tilde{y}_{i,T|T-1}$, are obtained for all N disaggregates using (10). Contemporaneous information is used to check the forecasts by testing for the significance of the impulse dated T in the J known

disaggregates using a tight significance level. If there is evidence of a break, incorporating the information into the forecasts for the $N - J$ unknown disaggregates will depend on judgment as to whether the break is also expected to occur in other disaggregate series in the same block. This will depend on: (a) whether the significant outlier is assumed to be measurement error or a location shift; and (b) whether is assumed to be correlated with the unknown disaggregates at T . Historical data will only be partly informative as the correlations between disaggregates usually also change when there is a location shift: see Castle *et al.* (2009c). Hence, we formalize the role of judgment using the forecasting rule:

$$\hat{y}_{i,T|T} = (1 - l_k) \tilde{y}_{i,T|T-1} + l_k \tilde{y}_{i,T|T}^*, \quad i = J + 1, \dots, N, \quad (23)$$

where l_k is an indicator function taking the value 1 when $k\alpha J \geq p$ for a small integer k where $p = \sum_{i=1}^J 1_T$ is the number of impulses retained at T for the J known disaggregates, and $\tilde{y}_{i,T|T}^*$ is an intercept corrected forecast given by:

$$\tilde{y}_{i,T|T}^* = \tilde{y}_{i,T|T-1} + \frac{1}{p} \sum_{j=1}^p \hat{\varsigma}_{p,T} \quad (24)$$

This uses an average of the retained impulses, but a weighted average could be used based on correlations between the disaggregate series.

Clements and Hendry (1998, ch.8) outline a theory of intercept corrections. Illustrating by the simplest DGP for the disaggregates:

$$y_{i,T} = \psi_i y_{i,T-1} + v_{i,T} \quad (25)$$

then when $l_k = 0$:

$$\tilde{y}_{i,T|T-1} = \hat{\psi}_i y_{i,T-1}$$

and when $l_k = 1$:

$$\tilde{y}_{i,T|T}^* = \tilde{y}_{i,T|T-1} + \mu_{p,T}$$

where $\mu_{p,T} = \sum_{j=1}^p \hat{\varsigma}_{p,T}$, so the forecast error is given by:

$$\begin{aligned} \tilde{v}_{i,T}^* &= y_{i,T} - \tilde{y}_{i,T|T}^* \\ &= \left(\psi_i - \hat{\psi}_i \right) y_{i,T-1} + v_{i,T} - \mu_{p,T} \\ &= \tilde{v}_{i,T} - \mu_{p,T} \end{aligned} \quad (26)$$

where $\tilde{v}_{i,T}$ is the forecast error from the uncorrected forecast. (24) is analogous to setting the forecast back on track but across disaggregate series as opposed to through time, hence nowcasting as opposed to forecasting. This implies that the error variance, $V[\tilde{v}_{i,T}^*]$ is essentially equal to that of the uncorrected error variance, $V[\tilde{v}_{i,T}]$ as $\mu_{p,T}$ is nearly a fixed constant for each i when J is large. This is in contrast to a standard intercept correction that sets the forecast back on track and doubles the error variance.

Hence, if a number of known disaggregates break at T , then the forecasts for the unknown disaggregates should be adjusted; otherwise the regular forecasts are used. In practice, the

disaggregates should be split into blocks corresponding to groups of variables such as industrial production, prices, financial variables, interest rates, labour variables, housing market variables, etc. Ideally, the blocks would capture sets of variables with close linkages, such that a break in one series is likely to spread through to the other series. The blocks also need to combine releases with different timings, such that the strategy of using contemporaneous information could be applied within block. The blocks can be determined by the presence of common trends or cycles for subsets of disaggregates captured by factors estimated using static principal components as discussed above. Blocks of disaggregates that exhibit common trends or cycles are more likely to be susceptible to common breaks, although this is by no means guaranteed. p would refer to the number of series with significant impulses at T within each block, so the intercept correction is applied to each block separately. *Ex post*, the cause of the break should be investigated when resources permit, but this may not be feasible in a nowcasting context.

7.1 Differencing

After the break has occurred, if there is evidence that the break will persist, so a location shift has occurred, this will be detectable from the J observed disaggregates at $T + 1$. The differenced VEqCM forecast discussed in section 6.1 will be robust once the break has occurred, so $\tilde{y}_{i,T+1|T}^*$ should be based on the differenced model (17) or (19). Unlike intercept correction, this method does not impose the magnitude of the break on the disaggregate forecast, but the forecast will be robust to breaks after the break has occurred. Two observations are required at a minimum to determine whether the break is a permanent location shift or a transitory break. As data for the J known disaggregates are available at T , once observation $T + 1$ is available, those two observations can help to distinguish between the two hypotheses, enabling robust nowcasts for the unknown disaggregates one period after the break. This requires that the same J set of components are available at T and $T + 1$. Otherwise evidence is only available at $T + 2$ and judgment is required to determine whether the break is thought to be permanent or transitory at T and $T + 1$.

8 Conclusion

Nowcasting is an essential component of macroeconomic policy analysis. Current-dated estimates form the inputs for longer term forecasting, so a reconstruction of incomplete current data are almost always required. All forms of forecasting rely to some extent on judgment, equally so for nowcasting. We have proposed a computationally feasible method for doing so, allowing for structural breaks of the location shift form both in-sample and at the forecast origin, covariate information and a subset of observed disaggregates, using automatic model selection methods that can handle more variables than observations. The approach of forecasting the disaggregate series and aggregating uses a larger information set than forecasting the aggregate directly and enables rapid detection of breaks, switching to robust forecasting methods when breaks are detected.

However, there are caveats to the proposed approach. There is a signal extraction problem as

structural breaks and measurement errors at the forecast origin are observationally equivalent, an issue that remains to be resolved. A large cross-section dimension and higher-frequency data can help to disentangle the two explanations, and we propose a conservative strategy for detecting breaks, so the forecasts will only be adapted if there is strong evidence that a break has occurred. The approach relies on there being some proportion of the disaggregates, J , that are timely and contain a signal about the $N - J$ unknown disaggregates, but this seems a reasonable assumption given the high correlations between disaggregates, especially within the blocks sharing common features estimated by principal component analysis. Finally, data are released at varying times throughout the month, but this can be accommodated in our approach, so the method ensures that the largest available information set is used at each nowcast origin. The proposed strategy should be operationally feasible in a real-time context but awaits future empirical applications.

References

- Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica*, **61**, 821–856.
- Bai, J., and Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, **66**, 47–78.
- Castle, J. L., Doornik, J. A., and Hendry, D. F. (2009a). Model selection when there are multiple breaks. Working paper, Economics Department, University of Oxford.
- Castle, J. L., Doornik, J. A., Hendry, D. F., and Nymoen, R. (2009b). Testing the invariance of expectations models of inflation. Working paper, Economics Department, University of Oxford.
- Castle, J. L., Fawcett, N. W. P., and Hendry, D. F. (2009c). Forecasting with equilibrium-correction models during structural breaks. *Journal of Econometrics*, **forthcoming**, –.
- Castle, J. L., Fawcett, N. W. P., and Hendry, D. F. (2009d). Nowcasting as contemporaneous forecasting. Mimeo, Economics Department, Oxford University.
- Castle, J. L., and Shephard, N. (eds.)(2009). *The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry*. Oxford: Oxford University Press.
- Clements, M. P., and Hendry, D. F. (1998). *Forecasting Economic Time Series*. Cambridge: Cambridge University Press.
- Clements, M. P., and Hendry, D. F. (1999). *Forecasting Non-stationary Economic Time Series*. Cambridge, Mass.: MIT Press.
- Clements, M. P., and Hendry, D. F. (2001). Explaining the results of the M3 forecasting competition. *International Journal of Forecasting*, **17**, 550–554.
- Clements, M. P., and Hendry, D. F. (2003). Forecasting in the National Accounts at the Office for National Statistics. Report no 12, Statistics Commission.
- Clements, M. P., and Hendry, D. F. (2005). Guest Editors’ introduction: Information in economic forecasting. *Oxford Bulletin of Economics and Statistics*, **67**, 713–753.

- Clements, M. P., and Hendry, D. F. (2008). Economic forecasting in a changing world. *Capitalism and Society*, **3**, 1–18.
- Doornik, J. A. (2009). Autometrics. in Castle, and Shephard (2009), pp. 88–121.
- Ferrara, L., Guegan, D., and Rakotomalahy, P. (2008). Gdp nowcasting with ragged-edge data: A semi-parametric modelling. Ces working paper, 82, Paris School of Economics.
- Fildes, R., and Ord, K. (2002). Forecasting competitions – their role in improving forecasting practice and research. In Clements, M. P., and Hendry, D. F. (eds.), *A Companion to Economic Forecasting*, pp. 322–353: Oxford: Blackwells.
- Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2005). The generalized dynamic factor model: One-sided estimation and forecasting. *Journal of the American Statistical Association*, **100**, 830–840.
- Frisch, R., and Waugh, F. V. (1933). Partial time regression as compared with individual trends. *Econometrica*, **1**, 221–223.
- Giannone, D., Reichlin, L., and Small, D. (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics*, **55**, 665–676.
- Hendry, D. F. (2004). Forecasting long-run TV advertising expenditure in the UK. Commissioned report, Ofcom, London. <http://www.ofcom.org.uk/research/tv/reports/tvadvmarket.pdf>.
- Hendry, D. F. (2006). Robustifying forecasts from equilibrium-correction models. *Journal of Econometrics*, **135**, 399–426.
- 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 Clements, M. P. (2004). Pooling of forecasts. *Econometrics Journal*, **7**, 1–31.
- Hendry, D. F., and Hubrich, K. (2009). Combining disaggregate forecast or combining disaggregate information to forecast an aggregate. Working paper, Research Department, European Central Bank.
- 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 Reade, J. J. (2008). Modelling and forecasting using model averaging. Working paper, Economics Department, Oxford University.
- 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.
- Kuzin, V., Marcellino, M., and Schumacher, C. (2009). Pooling versus model selection for nowcasting with many predictors: An application to German GDP. Discussion paper series 1: Economic studies, no. 3, Deutsche Bundesbank.
- Lee, K., Olekalns, N., and Shields, K. (2008). Nowcasting, business cycle dating and the interpretation of new information when real time data are available. Discussion papers in economics 08/17, University of Leicester.

- Makridakis, S., and Hibon, M. (2000). The M3-competition: Results, conclusions and implications. *International Journal of Forecasting*, **16**, 451–476.
- Marcellino, M., and Schumacher, C. (2007). Factor-MIDAS for now- and forecasting with ragged-edge data: a model comparison for German GDP,. Discussion paper series 1: 34, Deutsche Bundesbank.
- Mayo, I., and Espasa, A. (2009). Forecasting aggregates and disaggregates with common features. Working paper, Universidad Carlos III, Madrid.
- Schumacher, C., and Breitung, J. (2008). Real-time forecasting of German GDP based on a large factor model with monthly and quarterly data. *International Journal of Forecasting*, **24**, 386–398.
- Stock, J. H., and Watson, M. W. (2002). Macroeconomic forecasting using diffusion indices. *Journal of Business and Economic Statistics*, **20**, 147–162.
- Wallis, K. (1986). Forecasting with an econometric model: The ‘ragged edge’ problem. *Journal of Forecasting*, **5**, 1–13.