Chapter 1
Open-model Forecast-error Taxonomies

David F. Hendry and Grayham E. Mizon*

Abstract
We develop forecast-error taxonomies when there are unmodeled variables, forecast 'off-line'. We establish three surprising results. Even when an open system is correctly specified in-sample with zero intercepts, despite known future values of strongly exogenous variables, changes in dynamics can induce forecast failure when they have non-zero means. The additional impact on forecast failure of incorrectly omitting such variables depends only on unanticipated shifts in their means. With no such shifts, there is no reduction in forecast failure from forecasting unmodeled variables relative to omitting them in 1-step or multi-step forecasts. Artificial data illustrations confirm these results.

JEL classifications: C51, C22.

Keywords: Forecasting; Forecast-error taxonomies; Location shifts; Open models.

1.1 Introduction

It is a pleasure to contribute a chapter on forecasting to a volume in honor of Hal White, as forecasting has long been a salient aspect of his research. We congratulate Hal on his major research findings and look forward to many more.

There are a number of taxonomies of the sources of forecast errors in closed systems where every variable to be forecast is modeled: see for example, [Clements and Hendry(1998), Clements and Hendry(2006)] and [Hendry and Hubrich(2011)]. Such taxonomies have clarified the problems facing forecasters when parameters change. Forecasting variables as part of systems that are subject to unanticipated changes is difficult, as recent floods, tsunamis, and the financial crisis demonstrate. Systematic forecast errors and forecast failures are mainly due to location shifts, namely changes in the previous unconditional means of the variables being forecast, and changes in other parameters can be hard to detect, as shown in [Hendry(2000)] and illustrated by [Hendry and Nielsen(2007)].

* Financial support from the Open Society Institute and the Oxford Martin School is gratefully acknowledged. We are indebted to Anindya Banerjee, Jennifer L. Castle, Mike Clements, Jurgen A. Doornik, Neil Ericsson, Katarina Juselius and John N.J. Muellbauer for helpful discussions about and comments on an earlier version.

In practice, many forecasting systems include unmodeled determinants, whose future values are determined 'off-line' by a separate process: examples include commodity prices, exchange rates, and outputs of trading partners. There are many reasons for not modeling some variables, namely those that are exceptionally difficult to forecast accurately, other variables that are policy instruments determined outside the system in use, and some weakly exogenous variables where conditioning on them incurs no loss of information for modeling (see [Engle et al(1983)Engle, Hendry, and Richard]). Using a taxonomy of the consequences of including or excluding 'off-line' variables as inputs in forecasting models, we clarify the forecasting problems which could result. Even when the forecasting model is correctly specified in-sample having accurately estimated coefficients, with unmodeled variables that are strongly exogenous and *known* into the future, nevertheless changes in the dynamics of the system can induce forecast failure simply because the unmodeled variables have non-zero means.

At first sight such a claim seems counter-intuitive: if a variable y_t is determined by

$$y_t = \gamma y_{t-1} + \lambda z_t + \varepsilon_t$$

say, when $\varepsilon_t \sim \mathsf{IN}[0, \sigma^2_\varepsilon]$, and z_t is strongly exogenous, then for known λz_t:

$$(y_t - \lambda z_t) = \gamma y_{t-1} + \varepsilon_t \tag{1.1}$$

where the right-hand side has no intercept. Hence it might seem that (1.1) is in the class of models where change is hard to detect. However, if z_t has a non-zero mean, then so does y_t, and that alone makes the model susceptible to forecast failure after any parameters change, and as we show below, that result holds whether or not z_t is included in the model.

There are four distinct scenarios to consider for 1-step ahead forecasts, when facing parameter shifts in the data-generation process (DGP). First, where strongly exogenous variables with known future values are correctly included in the forecasting model and all parameters are known (or sufficiently precisely estimated that sampling variation is a second-order issue). Second, when the strongly exogenous variables are unknowingly and incorrectly omitted. Third, when the strongly exogenous variables need to be forecast, either within the system or 'off-line'. Finally, allowing for parameter-estimation uncertainty, model mis-specification for the DGP and measurement errors at the forecast origin, a setting which in principle is applicable to all three previous cases but here is only considered for the third. An analogous four scenarios arise for multi-step forecasts, but as the key results seem little affected, we focus on the first four scenarios and briefly note extensions to forecasting more than one period ahead.

Section 1.2 investigates a correctly-specified $\mathsf{I}(0)$ open system to consider the sources of forecast failure that can result from changes in the parameters when the m unmodeled strongly exogenous variables, $\mathbf{z}_t$, have non-zero means. Section 1.2.1 investigates any additional impacts from unknowingly omitting the $\mathbf{z}_t$, and section 1.2.2 compares 1-step forecasts one period later in both those settings. Section 1.3 develops 1-step taxonomies, first for excluding the $\mathbf{z}_t$, then in §1.3.1 when they are forecast 'off-line', also allowing for parameter-estimation uncertainty, measurement errors at the forecast origin, and mis-forecasting the $\mathbf{z}_t$. Section 1.4 provides an artificial data illustration of the analytical results. Section 1.5 considers multi-step forecasts when the exogenous process is known in the future, then §1.5.1, §1.5.2 and §1.5.3 respectively consider the impacts of omitting the unmodeled variables, forecasting them, then parameter estimation. Section 1.6 briefly notes the transformations needed to reduce an initially $\mathsf{I}(1)$ system to $\mathsf{I}(0)$. Section 1.7 concludes. The appendix §1.8 compares forecasting in open and closed $\mathsf{I}(0)$ systems.

1.2 Forecasting in an open $\mathsf{I}(0)$ system

Consider an open $\mathsf{I}(0)$ system conditional on a set of m strongly exogenous variables $\{\mathbf{z}_t\}$,[2] which are known into the future (lagged unmodeled variables can be stacked within $\mathbf{z}_t$) where the conditional DGP over $t = 1, \ldots T$ is:

$$\mathbf{y}_t = \tau + \Upsilon y_{t-1} + \Gamma \mathbf{z}_t + \varepsilon_t \tag{1.2}$$

when $\varepsilon_t \sim \mathsf{IN}_n[\mathbf{0}, \Sigma]$ and $\mathsf{E}[\varepsilon_t|\mathbf{z}_1 \ldots \mathbf{z}_{T+H}] = \mathbf{0}$. A system which is $\mathsf{I}(1)$ and cointegrated is considered in §1.6. When all the variables are weakly stationary in-sample, so the eigenvalues of Υ lie within the unit circle, and we initially set all parameters to be constant, taking expectations in (1.2) when $\mathsf{E}[\mathbf{z}_t] = \rho$:

$$\mathsf{E}[\mathbf{y}_t] = \phi = \tau + \Upsilon\phi + \Gamma\mathsf{E}[\mathbf{z}_t] = \tau + \Upsilon\phi + \Gamma\rho,$$

so the in-sample equilibrium mean of $\mathbf{y}$ is:

$$\phi = (\mathbf{I}_n - \Upsilon)^{-1}(\tau + \Gamma\rho) \tag{1.3}$$

Consequently, we can re-write (1.2) as:

$$\mathbf{y}_t - \phi = \Upsilon(\mathbf{y}_{t-1} - \phi) + \Gamma(\mathbf{z}_t - \rho) + \varepsilon_t \tag{1.4}$$

Below, we use whichever parametrization (1.2) or (1.4) proves most convenient, although it must be remembered that how the means ϕ and ρ are connected in (1.3) depends on the invariants of the underlying behavior represented by agents' plans. For example, (1.3) only entails co-breaking between ϕ and ρ so long as the other parameters remain constant when ρ shifts (see e.g., [Hendry and Massmann(2007)], for an analysis of co-breaking). Concerning notation for forecast values, $\overline{\mathbf{y}}$ denotes a correctly specified model with known future $\mathbf{z}$; $\widetilde{\mathbf{y}}$ denotes when $\mathbf{z}$ is omitted from the model; and $\widehat{\mathbf{y}}$ is when the $\mathbf{z}$ are included in the model, but future values need to be forecast; and if needed, $\widehat{\widehat{\mathbf{y}}}$ for that last case when parameters are estimated. $\widehat{\mathbf{y}}_T$ denotes an estimated forecast-origin value.

We first consider a 1-step ahead forecast from time T for known $\mathbf{z}_{T+1}$ from a model that is correctly specified in-sample with known parameter values, denoted:

$$\overline{\mathbf{y}}_{T+1|T} = \tau + \Upsilon\mathbf{y}_T + \Gamma\mathbf{z}_{T+1} \tag{1.5}$$

However, the DGP in the next period in fact changes to:

$$\mathbf{y}_{T+1} = \tau^* + \Upsilon^*\mathbf{y}_T + \Gamma^*\mathbf{z}_{T+1} + \varepsilon_{T+1} \tag{1.6}$$

where all the parameters shift, including the dynamic feedback, and ρ shifts to $\mathsf{E}[\mathbf{z}_{T+1}] = \rho^*$. The resulting forecast error between (1.5) and (1.6) is $\overline{\varepsilon}_{T+1|T} = \mathbf{y}_{T+1} - \overline{\mathbf{y}}_{T+1|T}$ and hence:

$$\overline{\varepsilon}_{T+1|T} = (\tau^* - \tau) + (\Upsilon^* - \Upsilon)\mathbf{y}_T + (\Gamma^* - \Gamma) z_{T+1} + \varepsilon_{T+1} \tag{1.7}$$

so that:

[2] Corresponding to $\Psi_{zy} = \mathbf{0}$ in §1.8.

$$\begin{aligned}\mathsf{E}\left[\overline{\varepsilon}_{T+1|T}\right] &= (\tau^* - \tau) + (\Upsilon^* - \Upsilon)\,\mathsf{E}[\mathbf{y}_T] + (\Gamma^* - \Gamma)\mathsf{E}[\mathbf{z}_{T+1}] \\ &= (\tau^* - \tau) + (\Upsilon^* - \Upsilon)\,\phi + (\Gamma^* - \Gamma)\,\rho^* \end{aligned} \tag{1.8}$$

Consequently, even if $\tau^* = \tau = \mathbf{0}$ so (1.7) has no intercept and $\Gamma^* = \Gamma$ and $\rho^* = \rho$, so (1.8) then does not depend directly on $\mathbf{z}_{T+1}$ which anyway has constant parameters, nevertheless forecast failure can occur for $\rho \neq \mathbf{0}$ when $\Upsilon^* \neq \Upsilon$ as then:

$$\mathsf{E}\left[\overline{\varepsilon}_{T+1|T}\right] = (\Upsilon^* - \Upsilon)\,(\mathbf{I}_n - \Upsilon)^{-1}\,\Gamma\rho \tag{1.9}$$

which reveals an equilibrium-mean shift occurs in $\{\mathbf{y}_t\}$.

This outcome may be clearer when (1.4) is written using (1.3) as:

$$\mathbf{y}_t = (\mathbf{I}_n - \Upsilon)^{-1}\,(\tau + \Gamma\rho) + \Upsilon\,(\mathbf{y}_{t-1} - \phi) + \Gamma\,(\mathbf{z}_t - \rho) + \varepsilon_t \tag{1.10}$$

so that even when $\tau = \mathbf{0}$, although $(\mathbf{y}_{t-1} - \phi)$, $\Gamma\,(\mathbf{z}_t - \rho)$ and ε_t all have expectations of zero, (1.10) entails an equilibrium mean of

$$(\mathbf{I}_n - \Upsilon)^{-1}\,\Gamma\rho \tag{1.11}$$

which is zero only if $\rho = \mathbf{0}$ when $\Gamma \neq \mathbf{0}$.

This is our first main result: despite correctly including unmodeled strongly exogenous variables $\mathbf{z}_t$ with known future values in a forecasting equation with no intercept and known parameters, a change in dynamics alone can induce forecast failure when the $\mathbf{z}_t$ have non-zero means.

More surprising still is that such failure is little different to that resulting either from modeling and forecasting $\mathbf{z}_t$ (see §1.3) possibly by a vector autoregression (VAR) say, or even excluding $\mathbf{z}_t$ entirely from the model, either deliberately or inadvertently, as we now show in §1.2.1.

1.2.1 Omitting the exogenous variables

If it is not known that $\mathbf{z}_t$ is relevant, so it is inadvertently omitted, the mis-specified model of (1.4) is:

$$\mathbf{y}_t = \phi + \Upsilon_e\,(\mathbf{y}_{t-1} - \phi) + \mathbf{u}_t \tag{1.12}$$

where the subscript $_e$ in (1.12) denotes the finite-sample expected value following mis-specification (i.e., $\mathsf{E}[\widetilde{\Upsilon}] = \Upsilon_e$). Then $\mathbf{u}_t = \Gamma_e\,(\mathbf{z}_t - \rho) + \varepsilon_t$ with $\mathsf{E}[\mathbf{u}_t] = \mathbf{0}$. Provided there have not been any equilibrium-mean shifts in-sample, then $\phi_e = \phi$. The forecast using the expected parameter values (to abstract from sampling uncertainty) is:

$$\widetilde{\mathbf{y}}_{T+1|T} = \phi + \Upsilon_e\,(\mathbf{y}_T - \phi) \tag{1.13}$$

with $\widetilde{\mathbf{u}}_{T+1|T} = \mathbf{y}_{T+1} - \widetilde{\mathbf{y}}_{T+1|T}$ where (1.6) is reparametrized as:

$$\mathbf{y}_{T+1} = \phi^* + \Upsilon^*\,(\mathbf{y}_T - \phi^*) + \Gamma^*\,(\mathbf{z}_{T+1} - \rho^*) + \varepsilon_{T+1} \tag{1.14}$$

where $\phi^* = (\mathbf{I}_n - \Upsilon^*)^{-1}\,(\tau^* + \Gamma^*\rho^*)$. Then:

$$\begin{aligned}\widetilde{\mathbf{u}}_{T+1|T} &= (\phi^* - \phi) + \Upsilon^*\,(\mathbf{y}_T - \phi^*) - \Upsilon_e\,(\mathbf{y}_T - \phi) + \Gamma^*\,(\mathbf{z}_{T+1} - \rho^*) + \varepsilon_{T+1} \\ &= (\mathbf{I}_n - \Upsilon^*)\,(\phi^* - \phi) + (\Upsilon^* - \Upsilon_e)\,(\mathbf{y}_T - \phi) + \Gamma^*\,(\mathbf{z}_{T+1} - \rho^*) + \varepsilon_{T+1}\end{aligned} \tag{1.15}$$

with:

$$\mathsf{E}\left[\widetilde{\mathbf{u}}_{T+1|T}\right] = (\mathbf{I}_n - \Upsilon^*)(\phi^* - \phi)$$
$$= (\tau^* - \tau) + (\Upsilon^* - \Upsilon)\phi + (\Gamma^* - \Gamma)\rho^* + \Gamma(\rho^* - \rho) \tag{1.16}$$

Thus, (1.16) and (1.8) only differ by $\Gamma(\rho^* - \rho)$, and hence are the same when $\rho^* = \rho$ despite the mis-specification. When also $\tau^* = \tau = \mathbf{0}$ and $\Gamma^* = \Gamma$, both are non-zero at the value in (1.9). However, the forecast-error variances will differ between (1.15) and (1.7), with the former being larger in general.

This is our second main result: the additional impact on forecast failure of incorrectly omitting strongly exogenous variables depends only on shifts in their means. Combining these first two results, as the comparison of (1.16) and (1.8) shows, when their means are constant at zero, then irrespective of whether or not these strongly exogenous variables are included in the forecasting system, they neither cause nor augment forecast failure.

1.2.2 1-step forecasts one period later

The analyses of forecasting one period after a break in [Clements and Hendry(2011)] show that results can be substantively altered because of the impacts of the breaks on later data. From (1.6):

$$\mathbf{y}_{T+2} = \tau^* + \Upsilon^*\mathbf{y}_{T+1} + \Gamma^* z_{T+2} + \varepsilon_{T+2} \tag{1.17}$$

so that as $\mathsf{E}[\mathbf{z}_{T+2}] = \rho^*$:

$$\mathsf{E}[\mathbf{y}_{T+2}] = \tau^* + \Upsilon^*\mathsf{E}[\mathbf{y}_{T+1}] + \Gamma^*\mathsf{E}[\mathbf{z}_{T+2}] = \left(\mathbf{I}_n - (\Upsilon^*)^2\right)\phi^* + (\Upsilon^*)^2\phi$$
$$= \phi^* - (\Upsilon^*)^2(\phi^* - \phi) \tag{1.18}$$

as $\phi^* = (\mathbf{I}_n - \Upsilon^*)^{-1}(\tau^* + \Gamma^*\rho^*)$ and $\mathsf{E}[\mathbf{y}_{T+1}] = \phi^* - \Upsilon^*(\phi^* - \phi)$. Forecasting from (1.5) updated one period, but still with in-sample known parameters, so:[3]

$$\overline{\mathbf{y}}_{T+2|T+1} = \tau + \Upsilon y_{T+1} + \Gamma\mathbf{z}_{T+2} \tag{1.19}$$

the resulting forecast error $\overline{\varepsilon}_{T+2|T+1} = \mathbf{y}_{T+2} - \overline{\mathbf{y}}_{T+2|T+1}$ is:

$$\begin{aligned}
\overline{\varepsilon}_{T+2|T+1} = {} & (\mathbf{I}_n + \Upsilon^* - \Upsilon)(\mathbf{I}_n - \Upsilon^*)(\phi^* - \phi) && (Ia)\\
& + (\Upsilon^* - \Upsilon)(\mathbf{y}_{T+1} - \mathsf{E}[\mathbf{y}_{T+1}]) && (IIa)\\
& + \varepsilon_{T+2} && (IIIa)\\
& + (\Gamma^* - \Gamma)(\mathbf{z}_{T+2} - \rho^*) && (IVa)\\
& - \Gamma(\rho^* - \rho) && (Va)
\end{aligned} \tag{1.20}$$

with

$$\mathsf{E}\left[\overline{\varepsilon}_{T+2|T+1}\right] = (\mathbf{I}_n + \Upsilon^* - \Upsilon)(\mathbf{I}_n - \Upsilon^*)(\phi^* - \phi) - \Gamma(\rho^* - \rho) \tag{1.21}$$

[3] Recursive or moving-windows updating will drive the forecasting system towards the robust device considered in §1.2.3.

Similarly, omitting the $\mathbf{z}_{T+2}$, so using:

$$\widetilde{\mathbf{y}}_{T+2|T+1} = \phi + \Upsilon_e (\mathbf{y}_{T+1} - \phi) \tag{1.22}$$

then as:

$$\mathbf{y}_{T+2} = \phi^* - (\Upsilon^*)^2 (\phi^* - \phi) + \Upsilon^* (\mathbf{y}_{T+1} - \mathsf{E}[\mathbf{y}_{T+1}]) + \Gamma^* (\mathbf{z}_{T+2} - \rho^*) + \varepsilon_{T+2} \tag{1.23}$$

the forecast error $\widetilde{\varepsilon}_{T+2|T+1} = \mathbf{y}_{T+2} - \widetilde{\mathbf{y}}_{T+2|T+1}$ is:

$$\begin{aligned} \widetilde{\varepsilon}_{T+2|T+1} = & (\mathbf{I}_n + \Upsilon^* - \Upsilon_e)(\mathbf{I}_n - \Upsilon^*)(\phi^* - \phi) && (Ib) \\ & + (\Upsilon^* - \Upsilon_e)(\mathbf{y}_{T+1} - \mathsf{E}[\mathbf{y}_{T+1}]) && (IIb) \\ & + \varepsilon_{T+2} && (IIIb) \\ & + \Gamma^* (\mathbf{z}_{T+2} - \rho^*) && (IVb) \end{aligned} \tag{1.24}$$

with:

$$\mathsf{E}\left[\widetilde{\varepsilon}_{T+2|T+1}\right] = (\mathbf{I}_n + \Upsilon^* - \Upsilon_e)(\mathbf{I}_n - \Upsilon^*)(\phi^* - \phi) \tag{1.25}$$

Consequently, unlike [Clements and Hendry(2011)], comparing (1.21) and (1.25) shows that there are no substantive changes compared to the baseline case here, and those two formulae are essentially the same when $\rho^* = \rho$.

1.2.3 Avoiding systematic forecast failure

One implication of §1.2.2 is that until the forecasting model is changed, systematic forecast failure will persist. Out of the many possible methods for updating a model by intercept corrections, modeling the break, recursive or moving window re-estimation and differencing, we only note the last here (see [Hendry(2006)]). In place of (1.22), consider simply using the first-difference forecast, $\Delta\widetilde{\mathbf{y}}_{T+2|T+1} = \mathbf{0}$:

$$\widetilde{\mathbf{y}}_{T+2|T+1} = \mathbf{y}_{T+1} = \phi^* + \Upsilon^* (\mathbf{y}_T - \phi^*) + \Gamma^* (\mathbf{z}_{T+1} - \rho^*) + \varepsilon_{T+1}$$

so that using (1.23), $\widetilde{\varepsilon}_{T+2|T+1} = \mathbf{y}_{T+2} - \widetilde{\mathbf{y}}_{T+2|T+1}$ is:

$$\begin{aligned} \widetilde{\varepsilon}_{T+2|T+1} &= \phi^* - (\Upsilon^*)^2 (\phi^* - \phi) + \Upsilon^* (\mathbf{y}_{T+1} - \mathsf{E}[\mathbf{y}_{T+1}]) + \Gamma^* (\mathbf{z}_{T+2} - \rho^*) + \varepsilon_{T+2} \\ &\quad - \phi^* + \Upsilon^* (\phi^* - \phi) - \Upsilon^* (\mathbf{y}_T - \phi) - \Gamma^* (\mathbf{z}_{T+1} - \rho^*) - \varepsilon_{T+1} \\ &= \Upsilon^* (\mathbf{I}_n - \Upsilon^*)(\phi^* - \phi) + \Upsilon^* (\mathbf{y}_{T+1} - \mathsf{E}[\mathbf{y}_{T+1}]) - \Upsilon^* (\mathbf{y}_T - \phi) + \Gamma^* \Delta \mathbf{z}_{T+2} + \Delta \varepsilon_{T+2} \end{aligned}$$

so:

$$\mathsf{E}\left[\widetilde{\varepsilon}_{T+2|T+1}\right] = \Upsilon^* (\mathbf{I}_n - \Upsilon^*)(\phi^* - \phi) \tag{1.26}$$

which considerably dampens the forecast-error bias relative to (1.20) and (1.24) (e.g., for a univariate y_t, then $\Upsilon^* (1 - \Upsilon^*) \leq 0.25$).

1.3 1-step taxonomies

We now also allow for parameter estimation uncertainty, the mis-specification of omitting $\mathbf{z}$, and possible mis-measurement at the forecast origin, so the forecast-period DGP remains (1.14), whereas the forecasting model becomes:

$$\widetilde{\mathbf{y}}_{T+1|T} = \widetilde{\phi} + \widetilde{\Upsilon}\left(\widetilde{\mathbf{y}}_T - \widetilde{\phi}\right) \tag{1.27}$$

The forecast error, $\widetilde{\varepsilon}_{T+1|T} = \mathbf{y}_{T+1} - \widetilde{\mathbf{y}}_{T+1|T}$ can be decomposed into eleven empirically-relevant sources when $\phi_e \neq \phi$:

$$\begin{aligned}
\widetilde{\varepsilon}_{T+1|T} = {} & (\mathbf{I}_n - \Upsilon^*)(\phi^* - \phi) && \text{[1] equilibrium-mean shift} \\
& + (\Upsilon^* - \Upsilon)(\mathbf{y}_T - \phi) && \text{[2] dynamic shift} \\
& + (\mathbf{I}_n - \Upsilon_e)(\phi - \phi_e) && \text{[3] equilibrium-mean mis-specification} \\
& + (\Upsilon - \Upsilon_e)(\mathbf{y}_T - \phi) && \text{[4] dynamic mis-specification} \\
& + (\mathbf{I}_n - \Upsilon_e)\left(\phi_e - \widetilde{\phi}\right) && \text{[5] equilibrium-mean estimation} \\
& + \left(\Upsilon_e - \widetilde{\Upsilon}\right)(\mathbf{y}_T - \phi) && \text{[6] dynamic estimation} \\
& + \Upsilon_e(\mathbf{y}_T - \widetilde{\mathbf{y}}_T) && \text{[7] forecast origin mis-measurement} \\
& + \left(\widetilde{\Upsilon} - \Upsilon_e\right)\left(\widetilde{\phi} - \phi\right) && \text{[8] estimation covariance} \\
& + \left(\widetilde{\Upsilon} - \Upsilon_e\right)(\mathbf{y}_T - \widetilde{\mathbf{y}}_T) && \text{[9] measurement covariance} \\
& + \varepsilon_{T+1} && \text{[10] innovation error} \\
& + \Gamma^*(\mathbf{z}_{T+1} - \rho^*) && \text{[11] omitted variables}
\end{aligned} \tag{1.28}$$

As with earlier taxonomies, terms in (1.28) can be divided into those with non-zero expected values that lead to forecast biases, namely [1] and possibly [3] and [7] (noting that [8] is $\mathsf{O}_p(T^{-1})$), and those with zero means that only affect forecast-error variances, namely all the other terms, noting that $\mathsf{E}[\mathbf{y}_T - \phi] = \mathsf{E}[\mathbf{z}_{T+1} - \rho^*] = \mathbf{0}$. Thus, despite estimating a mis-specified system with omitted variables:

$$\mathsf{E}\left[\widetilde{\varepsilon}_{T+1|T}\right] \approx (\mathbf{I}_n - \Upsilon^*)(\phi^* - \phi) + \Upsilon_e(\phi - \mathsf{E}[\widetilde{\mathbf{y}}_T])$$

which matches (1.16) when $\mathsf{E}[\widetilde{\mathbf{y}}_T] = \phi$.

This outcome could be compared directly with that from including the known $\mathbf{z}_t$ in estimation and forecasting by dropping line [10], stacking $\mathbf{x}'_t = (\mathbf{y}'_t\, \mathbf{z}'_t)$ and redefining parameters, estimates and variables accordingly. Indeed, when $\Gamma^* = \Gamma = \mathbf{0}$, (1.28) becomes the forecast-error taxonomy for a VAR.

1.3.1 Forecasting the unmodeled variables

However, the more interesting and realistic case is where $\mathbf{z}_{T+1}$ is known to be relevant and has to be forecast with its parameters estimated in (1.2), which we now consider via:

$$\widehat{\mathbf{y}}_{T+1|T} = \widehat{\phi} + \widehat{\Upsilon}\left(\widehat{\mathbf{y}}_T - \widehat{\phi}\right) + \widehat{\Gamma}\left(\widehat{\mathbf{z}}_{T+1} - \widehat{\rho}\right) \tag{1.29}$$

compared to (1.6). Although the following derivation is under the correct specification of (1.29), the results above show that mis-specification does not create important additional problems, and for the dynamics, is already reflected in (1.28). Then, letting $\widehat{\varepsilon}_{T+1|T} = \mathbf{y}_{T+1} - \widehat{\mathbf{y}}_{T+1|T}$, all the terms from (1.28) remain other than [11] (still allowing for finite-sample biases in the dynamics, so $\Upsilon_e \neq \Upsilon$, but for simplicity taking $\mathsf{E}[\widehat{\rho}] = \rho$ and $\mathsf{E}[\widehat{\Gamma}] \approx \Gamma$) with the following 9 terms replacing the old [11].

$$\begin{aligned}
\widehat{\varepsilon}_{T+1|T} = \quad & && \text{[1]–[10] in (1.28)}\\
& -\Gamma\left(\rho^* - \rho\right) && \text{[11] exogenous mean shift}\\
& +\left(\Gamma^* - \Gamma\right)\left(\mathbf{z}_{T+1} - \rho^*\right) && \text{[12] exogenous slope shift}\\
& +\Gamma\left(\widehat{\rho} - \rho\right) && \text{[13] exogenous-mean estimation}\\
& -\left(\widehat{\Gamma} - \Gamma\right)\left(\mathbf{z}_{T+1} - \rho^*\right) && \text{[14] exogenous slope estimation}\\
& +\Gamma\left(\mathbf{z}_{T+1} - \mathsf{E}\left[\widehat{\mathbf{z}}_{T+1}\right]\right) && \text{[15] exogenous mean mis-forecast}\\
& +\left(\widehat{\Gamma} - \Gamma\right)\left(\widehat{\rho} - \rho\right) && \text{[16] estimation covariance}\\
& -\left(\widehat{\Gamma} - \Gamma\right)\left(\rho^* - \rho\right) && \text{[17] exogenous mean shift covariance}\\
& +\Gamma\left(\mathsf{E}\left[\widehat{\mathbf{z}}_{T+1}\right] - \widehat{\mathbf{z}}_{T+1}\right) && \text{[18] exogenous mis-forecast.}\\
& +\left(\widehat{\Gamma} - \Gamma\right)\left(\mathbf{z}_{T+1} - \widehat{\mathbf{z}}_{T+1}\right) && \text{[19] exogenous mis-forecast covariance}
\end{aligned} \tag{1.30}$$

We focus on the terms in (1.30) with non-zero expectations, where $\mathsf{E}\left[\mathbf{z}_{T+1}\right] = \rho^*$, and for simplicity covariances are ignored as a smaller order of magnitude. Then combined with (1.28):

$$\begin{aligned}
\mathsf{E}\left[\widehat{\varepsilon}_{T+1|T}\right] &\approx \left(\mathbf{I}_n - \Upsilon^*\right)\left(\phi^* - \phi\right) - \Gamma\left(\rho^* - \rho\right) + \Upsilon_e\left(\phi - \mathsf{E}\left[\widetilde{\mathbf{y}}_T\right]\right) + \Gamma\left(\rho^* - \mathsf{E}\left[\widehat{\mathbf{z}}_{T+1}\right]\right)\\
&= \left(\tau^* - \tau\right) + \left(\Upsilon^* - \Upsilon\right)\phi + \left(\Gamma^* - \Gamma\right)\rho^* + \Upsilon_e\left(\phi - \mathsf{E}\left[\widetilde{\mathbf{y}}_T\right]\right) + \Gamma\left(\rho^* - \mathsf{E}\left[\widehat{\mathbf{z}}_{T+1}\right]\right)
\end{aligned}$$

from (1.16). As before, when $\tau^* = \tau = \mathbf{0}$ and $\Gamma^* = \Gamma$, with $\mathsf{E}\left[\widetilde{\mathbf{y}}_T\right] = \phi$:

$$\mathsf{E}\left[\widehat{\varepsilon}_{T+1|T}\right] \approx \left(\Upsilon^* - \Upsilon\right)\phi + \Gamma\left(\rho^* - \mathsf{E}\left[\widehat{\mathbf{z}}_{T+1}\right]\right)$$

compared to $\mathsf{E}\left[\widetilde{\mathbf{u}}_{T+1|T}\right] = \left(\Upsilon^* - \Upsilon\right)\phi + \Gamma\left(\rho^* - \rho\right)$, so:

$$\mathsf{E}\left[\widehat{\varepsilon}_{T+1|T}\right] - \mathsf{E}\left[\widetilde{\mathbf{u}}_{T+1|T}\right] \approx -\Gamma\left(\mathsf{E}\left[\widehat{\mathbf{z}}_{T+1}\right] - \rho\right) \tag{1.31}$$

This is our third main result: exogenous variable forecasts have to be closer to the new mean ρ^* than the old mean ρ to deliver a smaller forecast-error bias than arises from omitting them.

When $\rho^* = \rho$, $\mathsf{E}\left[\widehat{\mathbf{z}}_{T+1}\right] = \rho$ is necessary for $\mathsf{E}\left[\widehat{\varepsilon}_{T+1|T}\right] = \mathsf{E}\left[\widetilde{\mathbf{u}}_{T+1|T}\right]$, and even then there will be variance effects both from parameter estimation and $\left(\mathsf{E}\left[\widehat{\mathbf{z}}_{T+1}\right] - \widehat{\mathbf{z}}_{T+1}\right)$. *This is our fourth main result*: when $\rho^* = \rho$, there is no reduction in forecast failure from accurately forecasting the exogenous variables relative to omitting them.

Our fifth main result is: this outcome does not depend on the strong exogeneity of the unmodeled variables, and holds even when they are only weakly exogenous.

Without strong assumptions about the dependencies between the many mean-zero terms in the taxonomy, it is not possible to derive explicit forecast-error variances, but it is clear there are many contributions beyond the innovation error variance, some of which could well be $\mathsf{O}_p(1)$, such as mis-forecasting the unmodeled variables, and forecast-origin mis-measurement. Moreover, as forecast errors could arise from every possible (non-repetitive) selection from the 19 terms, namely $\sum_{k=1}^{19} 19!/(19-k)! \approx 3.3 \times 10^{17}$, delineating their source must be nearly impossible.

1.4 Artificial data illustration

We consider a bivariate system with one unmodeled (strongly exogenous) variable, with known future values, where the baseline parameter values are $\tau = \mathbf{0}$ and $\rho = 0$ when:

$$\Upsilon = \begin{pmatrix} 0.5 & 0 \\ 0 & 0.5 \end{pmatrix} \;\; \Gamma = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \tag{1.32}$$

with $\Sigma = \mathbf{I}_2$, $T = 100$, and $h = 1, \ldots, 5$ one-step ahead forecasts after the break. The parameter shift investigated is:

$$\Upsilon^* = \begin{pmatrix} 0.75 & 0 \\ 0 & 0.5 \end{pmatrix} \tag{1.33}$$

first for the baseline, then when $\rho = 0$ but:

$$\tau = \begin{pmatrix} 5 \\ 0 \end{pmatrix} \tag{1.34}$$

and finally when $\tau = \mathbf{0}$ but $\rho = 5$.

The two equations are decoupled in this first experiment, whereas in the second:

$$\Upsilon = \begin{pmatrix} 0.5 & 0.5 \\ -0.3 & 0.5 \end{pmatrix} \tag{1.35}$$

again for the same scenarios.

The results of the first set are reported in Figure 1.1.

Panel a records forecasts $\widehat{y}_{1,T+h|T+h-1}$ from a single draw of the initial process in (1.32) when parameters are estimated, shown with error bands of $\pm 2\widehat{\sigma}_{11}$, and when including parameter estimation uncertainty, shown with bars. There is a very small increase in forecast uncertainty from adding parameter variances, consistent with an $\mathsf{O}_p(1/T)$ effect.

Panel b reports forecasts when z_t is omitted both in estimation and forecasting. Although the forecast intervals are wider, forecasts are similar and remain within their *ex ante* forecast intervals.

Panel c is for the correct specification but after the shift in (1.33), still with $\tau = \mathbf{0}$ and $\rho = 0$. Despite the break in the dynamics, forecasts remain within their *ex ante* forecast intervals, even though those are now incorrect. *Panel d* augments the problem by the incorrect omission of z_t, but hardly differs from *Panel b*.

Although we do not report the outcomes for a constant model and non-zero τ, they are well behaved around the new data outcomes. The same cannot be said for the outcomes in *Panels e & f* for the non-zero value of τ in (1.34) after the break in Υ in (1.33): forecast failure is manifest, and almost unaffected by whether z_t is included or omitted.

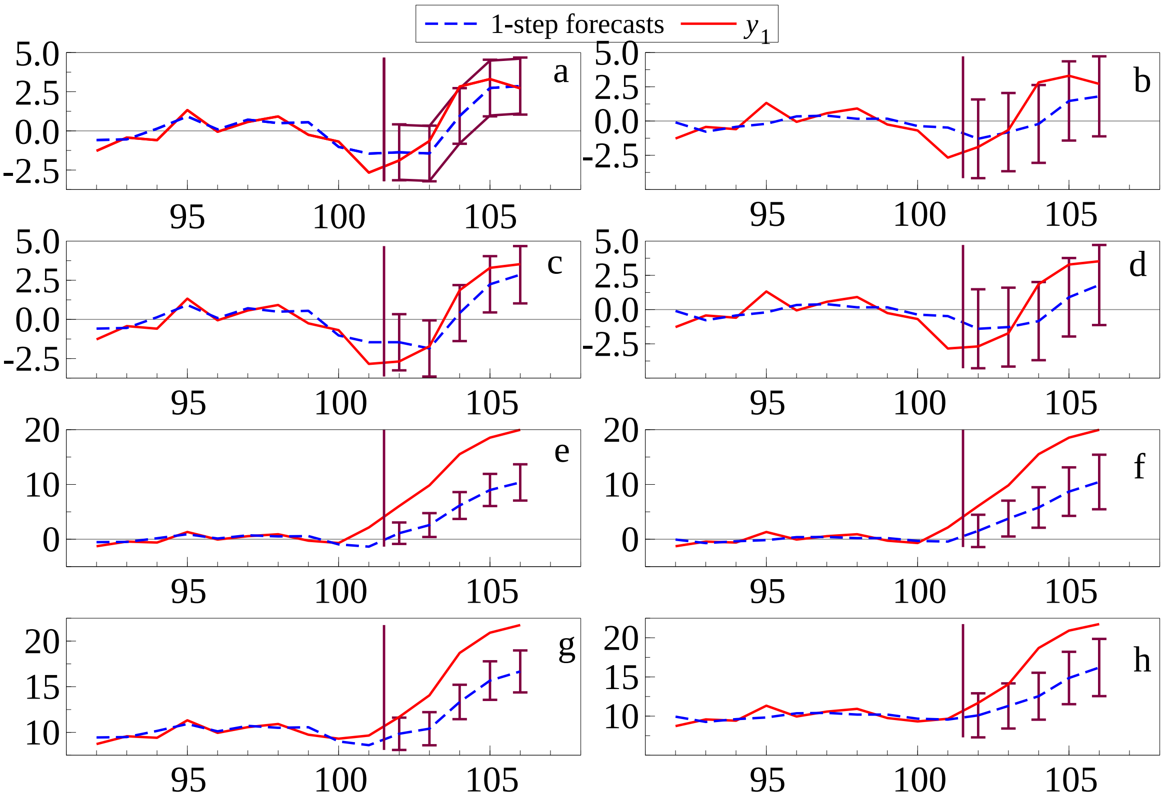

Fig. 1.1 Forecast failure for correct and mis-specified models

Finally, for $\rho = 5$, *Panels g & h* show the forecasts for the same break when the model is correctly specified by including z_t, and incorrect by omitting it. Despite the known future values of z_t and the absence of forecast failure after the break when $\rho = 0$, failure is again manifest and similar to *Panels e & f*.

The second setting in (1.35) yielded similar results, even though throughout both sets of experiments, the second variable was correctly forecast. All these results are consistent with the implications of the taxonomy in (1.28).

1.5 *h*-step ahead forecasts

We now consider the outcomes when an investigator needs to forecast h-steps ahead, $h > 1$. As the impacts of parameter-estimation uncertainty, mis-forecasting the unmodeled variables, and forecast-origin mis-measurement are similar to those derived above, we first derive the outcomes for known parameters to highlight the impacts of breaks when there are unmodeled variables. Thus, the in-sample system remains:

$$\mathbf{y}_t = \phi + \Upsilon(\mathbf{y}_{t-1} - \phi) + \Gamma(\mathbf{z}_t - \rho) + \epsilon_t$$

forecasting from $T+h-1$ to $T+h$ by:

$$\mathbf{y}_{T+h|T+h-1} = \phi + \Upsilon \left(\mathbf{y}_{T+h-1|T+h-2} - \phi\right) + \Gamma \left(\mathbf{z}_{T+h} - \rho\right)$$

leading to the multi-step forecast:

$$\mathbf{y}_{T+h|T} = \phi + \sum_{i=0}^{h-1} \Upsilon^i \Gamma \left(\mathbf{z}_{T+h-i} - \rho\right) + \Upsilon^h \left(\mathbf{y}_T - \phi\right) \tag{1.36}$$

If the system remained constant, the outcome would be:

$$\mathbf{y}_{T+h} = \phi + \sum_{i=0}^{h-1} \left[\Upsilon^i \Gamma \left(\mathbf{z}_{T+h-i} - \rho\right) + \Upsilon^i \varepsilon_{T+h-i}\right] + \Upsilon^h \left(\mathbf{y}_T - \phi\right) \tag{1.37}$$

so a known future $\{\mathbf{z}_t\}$ enters the same way as the cumulative error process. Then $\sum_{i=0}^{h-1} \Upsilon^i \varepsilon_{T+h-i}$ would be the only source of forecast error when equation (1.36) was used. However, that will not remain the case once there are changes in parameters, mis-specification of the model, or mis-estimation of Γ in (1.2), or unanticipated changes to ρ in the forecast period when the $\{\mathbf{z}_{T+h-i}\}$ are not known with certainty.

As before, we allow for structural change in the DGP, but to highlight the key problem, we first analyze a setting without estimation of, or mis-specification in, the econometrician's model for the DGP in-sample, so the in-sample parameter values are known. Under changes in all parameters of (1.37), the actual future outcomes will be:

$$\mathbf{y}_{T+h} = \phi^* + \sum_{i=0}^{h-1} \left(\Upsilon^*\right)^i \left[\Gamma^* \left(\mathbf{z}_{T+h-i} - \rho^*\right) + \varepsilon_{T+h-i}\right] + \left(\Upsilon^*\right)^h \left(\mathbf{y}_T - \phi^*\right) \tag{1.38}$$

When (1.36) is used, the forecast error $\mathbf{v}_{T+h|T} = \mathbf{y}_{T+h} - \mathbf{y}_{T+h|T}$ becomes:

$$\begin{aligned}
\mathbf{v}_{T+h|T} &= \phi^* - \phi + \left(\Upsilon^*\right)^h \left(\mathbf{y}_T - \phi^*\right) - \Upsilon^h \left(\mathbf{y}_T - \phi\right) \\
&\quad + \sum_{i=0}^{h-1} \left(\Upsilon^*\right)^i \Gamma^* \left(\mathbf{z}_{T+h-i} - \rho^*\right) - \sum_{i=0}^{h-1} \Upsilon^i \Gamma \left(\mathbf{z}_{T+h-i} - \rho\right) \\
&\quad + \sum_{i=0}^{h-1} \left(\Upsilon^*\right)^i \varepsilon_{T+h-i}
\end{aligned}$$

Taking these rows one at a time, and using:

$$\sum_{i=0}^{h-1} \mathbf{A}^i = \left(\mathbf{I}_n - \mathbf{A}^h\right) \left(\mathbf{I}_n - \mathbf{A}\right)^{-1}$$

first:

$$\begin{aligned}
&\phi^* - \phi + \left(\Upsilon^*\right)^h \left(\mathbf{y}_T - \phi^*\right) - \Upsilon^h \left(\mathbf{y}_T - \phi\right) \\
&= \left(\mathbf{I}_n - \left(\Upsilon^*\right)^h\right) \left(\phi^* - \phi\right) + \left(\left(\Upsilon^*\right)^h - \Upsilon^h\right) \left(\mathbf{y}_T - \phi\right)
\end{aligned}$$

where the terms respectively represent equilibrium-mean and slope shifts. Next:

$$\begin{aligned}
&\sum_{i=0}^{h-1}(\Upsilon^*)^i\Gamma^*(\mathbf{z}_{T+h-i}-\rho^*)-\sum_{i=0}^{h-1}\Upsilon^i\Gamma(\mathbf{z}_{T+h-i}-\rho)\\
&=\sum_{i=0}^{h-1}\left[(\Upsilon^*)^i\Gamma^*-\Upsilon^i\Gamma\right](\mathbf{z}_{T+h-i}-\rho^*)-\sum_{i=0}^{h-1}\Upsilon^i\Gamma(\rho^*-\rho)
\end{aligned}$$

where the first component has mean zero and the second is part of the exogenous mean shift. Finally, combining:

$$\begin{aligned}
\mathbf{v}_{T+h|T} &= \left(\mathbf{I}_n-(\Upsilon^*)^h\right)(\phi^*-\phi) && \text{[A] equilibrium-mean shift}\\
&+\left((\Upsilon^*)^h-\Upsilon^h\right)(\mathbf{y}_T-\phi) && \text{[B] dynamic shift}\\
&-\left(\mathbf{I}_n-\Upsilon^h\right)(\mathbf{I}_n-\Upsilon)^{-1}\Gamma(\rho^*-\rho) && \text{[C] exogenous mean shift}\\
&+\sum_{i=0}^{h-1}\left[(\Upsilon^*)^i\Gamma^*-\Upsilon^i\Gamma\right](\mathbf{z}_{T+h-i}-\rho^*) && \text{[D] exogenous slope shift}\\
&+\sum_{i=0}^{h-1}(\Upsilon^*)^i\epsilon_{T+h-i} && \text{[E] innovation error}
\end{aligned} \tag{1.39}$$

This outcome matches the earlier taxonomy specialized appropriately, namely [1], [2], [9], plus new [11] and [12]. As terms [A] and [C] have non-zero means, and the others have zero means:

$$\mathsf{E}[\mathbf{v}_{T+h}]=\left(\mathbf{I}_n-(\Upsilon^*)^h\right)(\phi^*-\phi)-\left(\mathbf{I}_n-\Upsilon^h\right)(\mathbf{I}_n-\Upsilon)^{-1}\Gamma(\rho^*-\rho) \tag{1.40}$$

Thus, even h-steps ahead, when $\rho^*=\rho$, forecast biases depend on $(\phi^*-\phi)$ which is non-zero whenever $\rho\neq\mathbf{0}$ despite $\tau^*=\tau=\mathbf{0}$.

This is our sixth main result: the first two results continue to hold for multi-step forecasts.

1.5.1 Omitting the unmodeled variables in h-step ahead forecasts

The forecasting model in-sample is now (1.4) leading to the multi-step forecasts:

$$\widetilde{\mathbf{y}}_{T+h|T}=\phi+(\Upsilon_e)^h(\mathbf{y}_T-\phi) \tag{1.41}$$

When (1.41) is used, the forecast error $\widetilde{\mathbf{v}}_{T+h|T}=\mathbf{y}_{T+h}-\widetilde{\mathbf{y}}_{T+h|T}$ becomes:

$$
\begin{aligned}
\widetilde{\mathbf{v}}_{T+h|T} &= (\phi^* - \phi) + (\Upsilon^*)^h (\mathbf{y}_T - \phi^*) - (\Upsilon_e)^h (\mathbf{y}_T - \phi) + \mathbf{v}_{T+h} \\
&= \left(\mathbf{I}_n - (\Upsilon^*)^h\right)(\phi^* - \phi) \\
&+ \left((\Upsilon^*)^h - (\Upsilon_e)^h\right)(\mathbf{y}_T - \phi) \\
&+ \sum_{i=0}^{h-1} (\Upsilon^*)^i \varepsilon_{T+h-i} \\
&+ \sum_{i=0}^{h-1} (\Upsilon^*)^i \Gamma^* (\mathbf{z}_{T+h-i} - \rho^*)
\end{aligned} \tag{1.42}
$$

matching the four terms in (1.15), where:

$$
\mathbf{v}_{T+h} = \sum_{i=0}^{h-1} (\Upsilon^*)^i \left[\Gamma^* (\mathbf{z}_{T+h-i} - \rho^*) + \varepsilon_{T+h-i}\right]
$$

with $\mathsf{E}[\mathbf{v}_{T+h}] = \mathbf{0}$, so that:

$$
\mathsf{E}\left[\widetilde{\mathbf{v}}_{T+h|T}\right] = \left(\mathbf{I}_n - (\Upsilon^*)^h\right)(\phi^* - \phi) \tag{1.43}
$$

This is our seventh main result: the previous conclusions about forecast failure based on the 1-step analyses are essentially unaltered: when $\rho^* = \rho$, (1.40) and (1.43) are equal, so forecast failure is only reduced by the inclusion of unmodeled variables when they have mean shifts.

1.5.2 Forecasting the unmodeled variables in h-step ahead forecasts

Now:

$$
\widehat{\mathbf{y}}_{T+h|T} = \phi + \sum_{i=0}^{h-1} \Upsilon^i \Gamma (\widehat{\mathbf{z}}_{T+h-i} - \rho) + \Upsilon^h (\mathbf{y}_T - \phi) \tag{1.44}
$$

with the forecast error $\widehat{\mathbf{v}}_{T+h|T} = \mathbf{y}_{T+h} - \widehat{\mathbf{y}}_{T+h|T}$:

$$
\begin{aligned}
\widehat{\mathbf{v}}_{T+h|T} &= (\phi^* - \phi) + (\Upsilon^*)^h (\mathbf{y}_T - \phi^*) - (\Upsilon_e)^h (\mathbf{y}_T - \phi) + \sum_{i=0}^{h-1} (\Upsilon^*)^i \varepsilon_{T+h-i} \\
&+ \sum_{i=0}^{h-1} (\Upsilon^*)^i \Gamma^* (\mathbf{z}_{T+h-i} - \rho^*) - \sum_{i=0}^{h-1} \Upsilon^i \Gamma (\widehat{\mathbf{z}}_{T+h-i} - \rho) \\
&= \left(\mathbf{I}_n - (\Upsilon^*)^h\right)(\phi^* - \phi) + \left((\Upsilon^*)^h - (\Upsilon_e)^h\right)(\mathbf{y}_T - \phi) + \sum_{i=0}^{h-1} (\Upsilon^*)^i \varepsilon_{T+h-i} \\
&+ \sum_{i=0}^{h-1} \left((\Upsilon^*)^i \Gamma^* - \Upsilon^i \Gamma\right)(\mathbf{z}_{T+h-i} - \rho^*) - \sum_{i=0}^{h-1} \Upsilon^i \Gamma (\widehat{\mathbf{z}}_{T+h-i} - \mathbf{z}_{T+h-i}) - \sum_{i=0}^{h-1} \Upsilon^i \Gamma (\rho^* - \rho)
\end{aligned} \tag{1.45}
$$

In the second block, the first three terms are identical to those in (1.42), and $\sum_{i=0}^{h-1}(\Upsilon^{*})^{i}\Gamma^{*}(\mathbf{z}_{T+h-i}-\rho^{*})$ has been replaced by terms relating to the shift in the dynamics (with mean zero), the forecast mistake, and the shift in the mean of the exogenous variables, as in (1.16).

1.5.3 Parameter estimation in h-step ahead forecasts

The estimated model forecasts are now:

$$\widehat{\widehat{\mathbf{y}}}_{T+h|T} = \widehat{\phi} + \sum_{i=0}^{h-1}\widehat{\Upsilon}^{i}\widehat{\Gamma}\left(\widehat{\mathbf{z}}_{T+h-i}-\widehat{\rho}\right) + \widehat{\Upsilon}^{h}\left(\widehat{\mathbf{y}}_{T}-\widehat{\phi}\right) \tag{1.46}$$

Thus, facing (1.38), the forecast error $\widehat{\widehat{\varepsilon}}_{T+h|T} = \mathbf{y}_{T+h} - \widehat{\widehat{\mathbf{y}}}_{T+h|T}$ is:

$$\begin{aligned}\widehat{\widehat{\varepsilon}}_{T+h|T} = {} & \phi^{*} - \widehat{\phi} + (\Upsilon^{*})^{h}(\mathbf{y}_{T}-\phi^{*}) - \widehat{\Upsilon}^{h}\left(\widehat{\mathbf{y}}_{T}-\widehat{\phi}\right) \\ & + \sum_{i=0}^{h-1}(\Upsilon^{*})^{i}\left[\Gamma^{*}(\mathbf{z}_{T+h-i}-\rho^{*})\right] - \sum_{i=0}^{h-1}\widehat{\Upsilon}^{i}\widehat{\Gamma}\left(\widehat{\mathbf{z}}_{T+h-i}-\widehat{\rho}\right) + \sum_{i=0}^{h-1}(\Upsilon^{*})^{i}\varepsilon_{T+h-i}\end{aligned}$$

which can be decomposed into the equivalent 19 terms as the earlier 1-step taxonomy in §1.3, analogous to the relation between (1.39) and (1.7). However, no new insights seem to be gained by doing so, and it is clear that the third result above still holds.

1.6 Transforming an $\mathsf{I}(1)$ system to $\mathsf{I}(0)$

Consider an n-dimensional $\mathsf{I}(1)$ VAR with p lags and an innovation error $\eta_t \sim \mathsf{IN}_n[\mathbf{0},\Omega_\eta]$ written as:

$$\mathbf{w}_t = \pi + \sum_{i=1}^{p}\Pi_i\mathbf{w}_{t-i} + \eta_t \tag{1.47}$$

where some of the np eigenvalues of the polynomial $\left|\mathbf{I}_n - \sum_{i=1}^{p}\Pi_i L^i\right|$ in L lie on, and the rest inside, the unit circle. Then $\Gamma = (\mathbf{I}_n - \sum_{i=1}^{p}\Pi_i)$ has reduced rank $0<r<n$, and can be expressed as $\Gamma = \alpha\beta'$ where α and β are $n\times r$ with rank r (see e.g., [Johansen(1995)]). Also $\pi = \gamma + \alpha\mu$, so when (e.g.) $p=2$:

$$\Delta\mathbf{w}_t = \gamma + (\Pi_1 - \mathbf{I}_n)(\Delta\mathbf{w}_{t-1}-\gamma) - \alpha\left(\beta'\mathbf{w}_{t-2}-\mu\right) + \eta_t \tag{1.48}$$

with $\mathsf{E}\left[\beta'\mathbf{w}_t\right] = \mu$ and $\mathsf{E}[\Delta\mathbf{w}_t] = \gamma$ where both $\Delta\mathbf{w}_t$ and $\beta'\mathbf{w}_t$ are $\mathsf{I}(0)$ even though $\mathbf{w}_t$ is $\mathsf{I}(1)$. Then r of the $\mathbf{x}_t$ above are $\beta'\mathbf{w}_t$ and $n-r$ are $\alpha'_{\perp}\Delta\mathbf{w}_t$ where $\alpha_{\perp}$ is $n\times(n-r)$ with $\alpha'_{\perp}\alpha = \mathbf{0}$ and $(\alpha:\alpha_{\perp})$ is non-singular.

Partitioning $\mathbf{w}_t$ into endogenous (modeled) variables $\mathbf{y}_t$ conditional on unmodeled $\mathbf{z}_t$ then produces an open system as analyzed in §1.8. Thus, our results hold in an open cointegrated system.

1.7 Conclusion

Even when a model is correctly specified in-sample, and the unmodeled variables, $\mathbf{z}_t$, are strongly exogenous with the correctly estimated coefficients, changes in the dynamics alone can induce forecast failure simply because the unmodeled variables have non-zero means. When the mean of $\mathbf{z}_t$ is constant, this forecast bias does not depend substantively on whether or not $\mathbf{z}_t$ is included in the forecasting model, but only on its non-zero mean. Including $\mathbf{z}_t$ in the forecasting model is beneficial when its mean shifts, but that advantage can be lost when future values $\mathbf{z}_{T+h}$ have to be forecast 'off-line'. These results are explicitly derived for one-step ahead forecasts and known parameters, but continue to hold when extended to estimated models, to multi-step forecasting, and to a later forecast origin following a break.

References

[Clements and Hendry(1998)] Clements MP, Hendry DF (1998) Forecasting Economic Time Series. Cambridge University Press, Cambridge

[Clements and Hendry(1999)] Clements MP, Hendry DF (1999) Forecasting Non-stationary Economic Time Series. MIT Press, Cambridge, Mass.

[Clements and Hendry(2006)] Clements MP, Hendry DF (2006) Forecasting with breaks. In: Elliott G, Granger CWJ, Timmermann A (eds) Handbook of Econometrics on Forecasting, Elsevier, Amsterdam, pp 605–657

[Clements and Hendry(2011)] Clements MP, Hendry DF (2011) Forecasting from Mis-specified Models in the Presence of Unanticipated Location Shifts. In: Clements MP, Hendry DF (eds) Oxford Handbook of Economic Forecasting, Oxford University Press, Oxford, pp 271–314

[Engle et al(1983)Engle, Hendry, and Richard] Engle RF, Hendry DF, Richard JF (1983) Exogeneity. Econometrica 51:277–304

[Hendry(2000)] Hendry DF (2000) On detectable and non-detectable structural change. Structural Change and Economic Dynamics 11:45–65

[Hendry(2006)] Hendry DF (2006) Robustifying forecasts from equilibrium-correction models. Journal of Econometrics 135:399–426

[Hendry and Hubrich(2011)] Hendry DF, Hubrich K (2011) Combining disaggregate forecasts or combining disaggregate information to forecast an aggregate. Journal of Business and Economic Statistics 29:216–227

[Hendry and Massmann(2007)] Hendry DF, Massmann M (2007) Co-breaking: Recent advances and a synopsis of the literature. Journal of Business and Economic Statistics 25:33–51

[Hendry and Nielsen(2007)] Hendry DF, Nielsen B (2007) Econometric Modeling: A Likelihood Approach. Princeton University Press, Princeton

[Johansen(1995)] Johansen S (1995) Likelihood-based Inference in Cointegrated Vector Autoregressive Models. Oxford University Press, Oxford

1.8 Appendix: comparing open with closed $\mathsf{I}(0)$ systems

Here, we relate the forecast-error taxonomy of the open conditional $\mathsf{I}(0)$ system in (1.2) to that for a closed VAR(1). Let $\mathbf{x}_t' = (\mathbf{y}_t' \, \mathbf{z}_t')$, then the DGP over $t = 1, \ldots, T$ for $\mathbf{y}_t$ and $\mathbf{z}_t$ is now:

$$\mathbf{x}_t = \psi + \Psi \mathbf{x}_{t-1} + \mathbf{v}_t \tag{1.49}$$

when $\mathbf{v}_t' = \left(\mathbf{v}_{yt}', \mathbf{v}_{zt}'\right) \sim \mathsf{IN}_{n+m}\left[\mathbf{0}, \boldsymbol{\Omega}\right]$ and $\boldsymbol{\Omega} = \begin{pmatrix} \boldsymbol{\Omega}_{yy} & \boldsymbol{\Omega}_{yz} \\ \boldsymbol{\Omega}_{zy} & \boldsymbol{\Omega}_{zz} \end{pmatrix}$ with $\boldsymbol{\Omega}_{zy} = \boldsymbol{\Omega}_{yz}'$. When all the variables are weakly stationary in-sample, taking expectations in (1.49):

$$\mathsf{E}\left[\mathbf{x}_t\right] = \boldsymbol{\psi} + \boldsymbol{\Psi}\mathsf{E}\left[\mathbf{x}_{t-1}\right] = \boldsymbol{\psi} + \boldsymbol{\Psi}\boldsymbol{\mu} = \boldsymbol{\mu},$$

so:

$$\boldsymbol{\mu} = \left(\mathbf{I}_{n+m} - \boldsymbol{\Psi}\right)^{-1}\boldsymbol{\psi} = \begin{pmatrix} \mathsf{E}\left[\mathbf{y}_t\right] \\ \mathsf{E}\left[\mathbf{z}_t\right] \end{pmatrix} = \begin{pmatrix} \boldsymbol{\phi} \\ \boldsymbol{\rho} \end{pmatrix}. \tag{1.50}$$

Consequently, we can re-write (1.49) as:

$$\mathbf{x}_t - \boldsymbol{\mu} = \boldsymbol{\Psi}\left(\mathbf{x}_{t-1} - \boldsymbol{\mu}\right) + \mathbf{v}_t \tag{1.51}$$

for $t = 1, 2, \ldots T$.

We first consider a one-step ahead forecast from time T from a model that is correctly specified in-sample with known parameter values:

$$\overline{\mathbf{x}}_{T+1|T} = \boldsymbol{\psi} + \boldsymbol{\Psi}\mathbf{x}_T \tag{1.52}$$

but where the DGP in the next period has shifted to:

$$\mathbf{x}_{T+1} = \boldsymbol{\psi}^* + \boldsymbol{\Psi}^*\mathbf{x}_T + \mathbf{v}_{T+1} \tag{1.53}$$

with $\mathbf{v}_{T+1} \sim \mathsf{IN}_{n+m}\left[\mathbf{0}, \boldsymbol{\Omega}\right]$. The resulting forecast error between (1.52) and (1.53) is $\overline{\mathbf{v}}_{T+1|T} = \mathbf{x}_{T+1} - \overline{\mathbf{x}}_{T+1|T}$ and hence:

$$\overline{\mathbf{v}}_{T+1|T} = \left(\boldsymbol{\psi}^* - \boldsymbol{\psi}\right) + \left(\boldsymbol{\Psi}^* - \boldsymbol{\Psi}\right)\mathbf{x}_T + \mathbf{v}_{T+1} \tag{1.54}$$

so that[4]:

$$\mathsf{E}\left[\overline{\mathbf{v}}_{T+1|T}\right] = \left(\boldsymbol{\psi}^* - \boldsymbol{\psi}\right) + \left(\boldsymbol{\Psi}^* - \boldsymbol{\Psi}\right)\mathsf{E}\left[\mathbf{x}_T\right] = \left(\boldsymbol{\mu}^* - \boldsymbol{\mu}\right). \tag{1.55}$$

From (1.50), we can partition (1.55) as:

$$\begin{aligned}
\mathsf{E}\left[\overline{\mathbf{v}}_{T+1|T}\right] &= \begin{pmatrix} \boldsymbol{\phi}^* - \boldsymbol{\phi} \\ \boldsymbol{\rho}^* - \boldsymbol{\rho} \end{pmatrix} \\
&= \begin{pmatrix} \left(\boldsymbol{\psi}_y^* - \boldsymbol{\psi}_y\right) \\ \left(\boldsymbol{\psi}_z^* - \boldsymbol{\psi}_z\right) \end{pmatrix} + \begin{pmatrix} \left(\boldsymbol{\Psi}_{yy}^* - \boldsymbol{\Psi}_{yy}\right) & \left(\boldsymbol{\Psi}_{yz}^* - \boldsymbol{\Psi}_{yz}\right) \\ \left(\boldsymbol{\Psi}_{zy}^* - \boldsymbol{\Psi}_{zy}\right) & \left(\boldsymbol{\Psi}_{zz}^* - \boldsymbol{\Psi}_{zz}\right) \end{pmatrix} \begin{pmatrix} \boldsymbol{\phi} \\ \boldsymbol{\rho} \end{pmatrix} \\
&= \begin{pmatrix} \nabla\boldsymbol{\psi}_y \\ \nabla\boldsymbol{\psi}_z \end{pmatrix} + \begin{pmatrix} \nabla\boldsymbol{\Psi}_{yy} & \nabla\boldsymbol{\Psi}_{yz} \\ \nabla\boldsymbol{\Psi}_{zy} & \nabla\boldsymbol{\Psi}_{zz} \end{pmatrix} \begin{pmatrix} \boldsymbol{\phi} \\ \boldsymbol{\rho} \end{pmatrix}
\end{aligned}$$

where ∇ denotes a change in a parameter, with:

$$\begin{array}{lll}
\nabla\boldsymbol{\psi}_y = \left(\boldsymbol{\psi}_y^* - \boldsymbol{\psi}_y\right) & \nabla\boldsymbol{\psi}_z = \left(\boldsymbol{\psi}_z^* - \boldsymbol{\psi}_z\right) & \nabla\boldsymbol{\Psi}_{yy} = \left(\boldsymbol{\Psi}_{yy}^* - \boldsymbol{\Psi}_{yy}\right) \\
\nabla\boldsymbol{\Psi}_{yz} = \left(\boldsymbol{\Psi}_{yz}^* - \boldsymbol{\Psi}_{yz}\right) & \nabla\boldsymbol{\Psi}_{zy} = \left(\boldsymbol{\Psi}_{zy}^* - \boldsymbol{\Psi}_{zy}\right) & \nabla\boldsymbol{\Psi}_{zz} = \left(\boldsymbol{\Psi}_{zz}^* - \blacksquare_{zz}\right)
\end{array}$$

Partitioning $\boldsymbol{\mu} = \left(\mathbf{I}_{n+m} - \boldsymbol{\Psi}\right)^{-1}\boldsymbol{\psi}$ yields:

[4] Note that although $\mathsf{E}\left[\mathbf{x}_t\right] = \boldsymbol{\psi} + \boldsymbol{\Psi}\mathsf{E}\left[\mathbf{x}_{t-1}\right] = \boldsymbol{\psi} + \boldsymbol{\Psi}\boldsymbol{\mu} = \boldsymbol{\mu}$ for $t = 1, 2, \ldots T$ when $t > T$ $\mathsf{E}\left[\mathbf{x}_{T+j}\right] = \sum_{i=0}^{j-1}\left(\boldsymbol{\Psi}^*\right)^i\boldsymbol{\psi}^* + \left(\boldsymbol{\Psi}^*\right)^j\boldsymbol{\mu}$ for $j \geqq 1$ which for a stationary $\{\mathbf{x}_t\}$ process converges to $\left(\mathbf{I}_{n+m} - \boldsymbol{\Psi}^*\right)^{-1}\boldsymbol{\psi}^* = \boldsymbol{\mu}^*$ as $j \to \infty$.

$$\begin{pmatrix} \phi \\ \rho \end{pmatrix} = \begin{pmatrix} (\mathbf{I}_n - \Psi_{yy}) & -\Psi_{yz} \\ -\Psi_{zy} & (\mathbf{I}_m - \Psi_{zz}) \end{pmatrix}^{-1} \begin{pmatrix} \psi_y \\ \psi_z \end{pmatrix}$$
$$= \begin{pmatrix} \mathbf{A}^{-1} & -\mathbf{A}^{-1}\Psi_{yz}(\mathbf{I}_m - \Psi_{zz})^{-1} \\ -(\mathbf{I}_m - \Psi_{zz})^{-1}\Psi_{zy}\mathbf{A}^{-1} & \mathbf{B} \end{pmatrix} \begin{pmatrix} \psi_y \\ \psi_z \end{pmatrix}$$
$$= \begin{pmatrix} \mathbf{A}^{-1}\psi_y - \mathbf{A}^{-1}\Psi_{yz}(\mathbf{I}_m - \Psi_{zz})^{-1}\psi_z \\ \mathbf{B}\psi_z - (\mathbf{I}_m - \Psi_{zz})^{-1}\Psi_{zy}\mathbf{A}^{-1}\psi_y \end{pmatrix} \tag{1.56}$$

when $\mathbf{A} = [(\mathbf{I}_n - \Psi_{yy}) - \Psi_{yz}(\mathbf{I}_m - \Psi_{zz})^{-1}\Psi_{zy}]$ and $\mathbf{B} = (\mathbf{I}_m - \Psi_{zz})^{-1}[\mathbf{I}_m + \Psi_{zy}\mathbf{A}^{-1}\Psi_{yz}(\mathbf{I}_m - \Psi_{zz})^{-1}]$. Therefore (1.55) has the form:

$$\mathsf{E}\left[\overline{\mathbf{v}}_{T+1|T}\right] = \begin{pmatrix} \nabla\psi_y \\ \nabla\psi_z \end{pmatrix} + \begin{pmatrix} \nabla\Psi_{yy} & \nabla\Psi_{yz} \\ \nabla\Psi_{zy} & \nabla\Psi_{zz} \end{pmatrix} \begin{pmatrix} \mathbf{A}^{-1}\psi_y - \mathbf{A}^{-1}\Psi_{yz}(\mathbf{I}_m - \Psi_{zz})^{-1}\psi_z \\ \mathbf{B}\psi_z - (\mathbf{I}_m - \Psi_{zz})^{-1}\Psi_{zy}\mathbf{A}^{-1}\psi_y \end{pmatrix}$$

Hence, if the mean of the $\{\mathbf{z}_t\}$ process is constant ($\nabla\psi_z = \mathbf{0}, \nabla\Psi_{zy} = \mathbf{0}, \nabla\Psi_{zz} = \mathbf{0}$), and there is no intercept in the $\{\mathbf{y}_t\}$ process ($\psi_y^* = \psi_y = \mathbf{0}$), the mean of the forecast error becomes:

$$\mathsf{E}\left[\overline{\mathbf{v}}_{T+1|T}\right] = \begin{pmatrix} \left\{\nabla\Psi_{yz}\mathbf{B} - \nabla\Psi_{yy}\mathbf{A}^{-1}\Psi_{yz}(\mathbf{I}_m - \Psi_{zz})^{-1}\right\}\psi_z \\ \mathbf{0} \end{pmatrix}$$

so if there is a change in the dynamics of the $\{\mathbf{y}_t\}$ process and $\{\mathbf{z}_t\}$ has a non-zero mean, there will be forecast failure. Further, even if $\mathbf{z}_t$ is strongly exogenous for the parameters of the $\{\mathbf{y}_t\}$ process ($\Psi_{zy} = \mathbf{0}$), there will be forecast failure as:

$$\mathsf{E}\left[\overline{\mathbf{v}}_{T+1|T}\right] = \begin{pmatrix} \left\{\nabla\Psi_{yz} - \nabla\Psi_{yy}\mathbf{A}^{-1}\Psi_{yz}\right\}(\mathbf{I}_m - \Psi_{zz})^{-1}\psi_z \\ \mathbf{0} \end{pmatrix}$$

which will be non-zero provided $\psi_z \neq \mathbf{0}$ and there is a change in the dynamics of the $\{\mathbf{y}_t\}$ process, consistent with the closed system results in [Clements and Hendry(1999)].

These closed system results can be mapped to an open system using a conditional and marginal factorization of the joint distribution. From (1.49), the conditional distribution of $\mathbf{y}_t$ given $\mathbf{z}_t$ and the past is:

$$\mathbf{y}_t = \left(\psi_y - \Xi\psi_z\right) + (\Psi_{yy} - \Xi\Psi_{zy})\,\mathbf{y}_{t-1} + \Xi\mathbf{z}_t + (\Psi_{yz} - \Xi\Psi_{zz})\,\mathbf{z}_{t-1} + (\mathbf{v}_{yt} + \Xi\mathbf{v}_{zt})$$
$$= \theta + \Theta y_{t-1} + \Xi\mathbf{z}_t + \Lambda z_{t-1} + \nu_t \tag{1.57}$$

when $\Xi = \Omega_{yz}\Omega_{zz}^{-1}$. The initial VAR formulation induces one lag longer in $\mathbf{z}_t$ with:

$$\mathsf{E}[\mathbf{y}_t] = \theta + \Theta\mathsf{E}[\mathbf{y}_{t-1}] + \Xi\mathsf{E}[\mathbf{z}_t] + \Lambda\mathsf{E}[\mathbf{z}_{t-1}] = \theta + \Theta\phi + (\Xi + \Lambda)\rho = \phi$$

so that:

$$\phi = (\mathbf{I}_n - \Theta)^{-1}\{\theta + (\Xi + \Lambda)\rho\}$$

and:

$$(\mathbf{y}_t - \phi) = \Theta(\mathbf{y}_{t-1} - \phi) + \Xi(\mathbf{z}_t - \rho) + \Lambda(\mathbf{z}_{t-1} - \rho) + \nu_t.$$

The forecast error from predicting $\mathbf{y}_{T+1}$ by $\overline{\mathbf{y}}_{T+1|T} = \theta + \Theta\mathbf{y}_{t-1} + \Xi\mathbf{z}_{T+1} + \Lambda\mathbf{z}_T$ with known parameters and $\mathbf{z}_{T+1}$ and $\mathbf{z}_T$ is:

$$\overline{\nu}_{T+1} = \mathbf{y}_{T+1} - \overline{\mathbf{y}}_{T+1|T} = \nabla\theta + \nabla\Theta\mathbf{y}_{t-1} + \nabla\Xi\mathbf{z}_{T+1} + \nabla\Lambda\mathbf{z}_T + \nu_{T+1}$$

hence:

$$\begin{aligned}\mathsf{E}\left[\overline{\nu}_{T+1}\right] &= \nabla\theta + \nabla\Theta\phi + (\nabla\Xi + \nabla\Lambda)\,\rho \\ &= \nabla\theta + \nabla\Theta\left(\mathbf{I}_n - \Theta\right)^{-1}\left\{\theta + (\Xi + \Lambda)\,\rho\right\} + (\nabla\Xi + \nabla\Lambda)\,\rho\end{aligned}$$

with

$$\begin{aligned}\rho &= \mathbf{B}\psi_z - (\mathbf{I}_m - \Psi_{zz})^{-1}\Psi_{zy}\mathbf{A}^{-1}\psi_y \\ \nabla\theta &= (\nabla\psi_y - \nabla\Xi\psi_z - \Xi\nabla\psi_z) \\ \nabla\Theta &= (\nabla\Psi_{yy} - \nabla\Xi\Psi_{zy} - \Xi\nabla\Psi_{zy}) \\ \nabla\Lambda &= (\nabla\Psi_{yz} - \nabla\Xi\Psi_{zz} - \Xi\nabla\Psi_{zz})\end{aligned}$$

If the $\{\mathbf{z}_t\}$ process is constant ($\nabla\psi_z = \mathbf{0}, \nabla\Psi_{zy} = \mathbf{0}, \nabla\Psi_{zz} = \mathbf{0}$) and there is no intercept in the $\{\mathbf{y}_t\}$ process ($\psi_y^* = \psi_y = \mathbf{0}$) then $\rho = \mathbf{B}\psi_z$ and the mean of the forecast error becomes:

$$\begin{aligned}\mathsf{E}\left[\overline{\nu}_{T+1}\right] = &-\left[\nabla\Xi + (\nabla\Psi_{yy} - \nabla\Xi\Psi_{zy})\,(\mathbf{I}_n - \Psi_{yy} + \Xi\Psi_{zy})^{-1}\,\Xi\right]\psi_z \\ &+ \left[(\nabla\Psi_{yy} - \nabla\Xi\Psi_{zy})\,(\mathbf{I}_n - \Psi_{yy} + \Xi\Psi_{zy})^{-1}\,(\Xi + \Lambda) + (\nabla\Xi + \nabla\Lambda)\right]\mathbf{B}\psi_z\end{aligned}$$

which, when $\mathbf{z}_t$ is strongly exogenous for the parameters of the $\{\mathbf{y}_t\}$ process ($\Psi_{zy} = \mathbf{0}$), simplifies to:

$$\begin{aligned}\mathsf{E}\left[\overline{\nu}_{T+1}\right] = &-\left[\nabla\Xi + \nabla\Psi_{yy}\,(\mathbf{I}_n - \Psi_{yy})^{-1}\,\Xi\right]\psi_z \\ &+ \left[\nabla\Psi_{yy}\,(\mathbf{I}_n - \Psi_{yy})^{-1}\,(\Xi + \Lambda) + (\nabla\Xi + \nabla\Lambda)\right](\mathbf{I}_m - \Psi_{zz})^{-1}\psi_z\end{aligned}$$

so again will be non-zero when $\psi_z \neq \mathbf{0}$ and there is a change in the dynamics of the $\{\mathbf{y}_t\}$ process (i.e., at least one of $\nabla\Psi_{yy}, \nabla\Xi$ and $\nabla\Lambda$ is non-zero). This result mirrors that in (1.8) noting that $\rho = (\mathbf{I}_m - \Psi_{zz})^{-1}\psi_z$ in this case.

An analogous result is obtained when we close the open conditional $\mathsf{I}(0)$ system in (1.2) by endogenizing $\mathbf{z}_t$ in:

$$\mathbf{y}_t = \tau + \Upsilon\mathbf{y}_{t-1} + \Gamma\mathbf{z}_t + \varepsilon_t \tag{1.58}$$

$$\mathbf{z}_t = \lambda + \Phi\mathbf{y}_{t-1} + \Pi\mathbf{z}_{t-1} + \eta_t \tag{1.59}$$

so that:

$$\mathsf{E}\left[\mathbf{z}_t\right] = \lambda + \Phi\mathsf{E}\left[\mathbf{y}_{t-1}\right] + \Pi\mathsf{E}\left[\mathbf{z}_{t-1}\right] = \lambda + \Phi\phi + \Pi\rho = \rho$$

or:

$$\lambda = (\mathbf{I}_m - \Pi)\,\rho - \Phi\phi$$

leading to:

$$(\mathbf{z}_t - \rho) = \Phi\,(\mathbf{y}_{t-1} - \phi) + \Pi\,(\mathbf{z}_{t-1} - \rho) + \eta_t$$

Then, as $\phi = (\mathbf{I}_n - \Upsilon)^{-1}\,(\tau + \Gamma\rho)$:[5]

[5] This is true whether or not $\mathbf{z}_t$ is strongly exogenous (i.e., $\Phi = \mathbf{0}$) for the parameters of $\mathbf{y}_t$ in the VAR.

$$\begin{aligned}\mathbf{y}_t - \phi &= \Upsilon(\mathbf{y}_{t-1} - \phi) + \Gamma(\mathbf{z}_t - \rho) + \varepsilon_t \\ &= (\Upsilon + \Gamma\Phi)(\mathbf{y}_{t-1} - \phi) + \Gamma\Pi(\mathbf{z}_{t-1} - \rho) + (\Gamma\eta_t + \varepsilon_t)\end{aligned}$$

These results allow a general evaluation of the relative impacts of breaks when $\mathbf{z}_t$ is treated as 'external' or 'internal'.