Analyzing Differences between Scenarios

David F. Hendry and Felix Pretis[*]
Nuffield College and Institute for New Economic Thinking at the
Oxford Martin School, University of Oxford, UK;
and University of Victoria, Canada.

Abstract

Comparisons between alternative scenarios are used in many disciplines from macroeconomics to climate science to help with planning future responses. Differences between scenario paths are often interpreted as signifying likely differences between outcomes that would materialise in reality. However, even when using correctly specified statistical models of the in-sample data generation process, additional conditions are needed to sustain inferences about differences between scenario paths. We consider two questions in scenario analyses: First, does testing the difference between scenarios yield additional insight beyond simple tests conducted on the model estimated in-sample? Second, when does the estimated scenario difference yield unbiased estimates of the true difference in outcomes? Answering the first question, we show that the calculation of uncertainties around scenario differences raises difficult issues since the underlying in-sample distributions are identical for both 'potential' outcomes when the reported paths are deterministic functions. Under these circumstances, a scenario comparison adds little beyond testing for the significance of the perturbed variable in the estimated model. Resolving the second question, when models include multiple covariates, inferences about scenario differences depend on the relationships between the conditioning variables, especially their invariance to the interventions. Tests for invariance based on automatic detection of structural breaks can help identify in-sample invariance of models to evaluate likely constancy in projected scenarios. Applications of scenario analyses to impacts on the UK's wage share from unemployment and agricultural growth from climate change illustrate the concepts.

1 Introduction

Comparisons between alternative scenarios are used in many disciplines from economics (see e.g., Weale, 2016), climatology, and environmental sciences (see e.g., Webster, Babiker, Mayer, Reilly, Harnisch, Sarofim, and Wang, 2002, Webster, Forest, Reilly, Babiker, Kicklighter, Mayer, Prinn, Sarofim, Sokolov, Stone, and Wang, 2003, Burke, Hsiang, and Miguel, 2015, Pretis, Schwarz, Tang, Haustein, and Allen, 2018), to business (see e.g., Flyvbjerg, Holm, and Buhl, 2005) to help with planning future responses. Many have been produced recently to try and envision the future impacts of the COVID-19 pandemic. The differences between scenario paths are often interpreted as signifying likely differences between outcomes that would materialize in reality should that scenario be implemented. We consider two central questions in scenario analyses when the underlying models are estimated:

1. Does testing the difference between scenarios yield additional insights beyond tests conducted on the in-sample estimated model? We show that the answer is 'no' for simple models that are linear in a deterministically-perturbed scenario variable.

2. When does the estimated scenario difference yield unbiased estimates of the true difference in outcomes? The answer depends on the validity of the estimated model, the relationships between its covariates, and its invariance to the scenario perturbations.

Even when using correctly-specified models of the in-sample data generation process, itself an unlikely possibility in most observational-data disciplines, additional conditions are needed to sustain inferences about differences between scenario paths. Moreover, the calculation of uncertainties around scenario differences raises the difficult issue that the underlying in-sample distributions are identical for both 'potential' outcomes, and the reported paths are often deterministic functions. There are also a number of different types of scenario calculation, from deterministic impulse or step shifts applied to either a 'given', or conditioning, variable or the initial conditions of the target variable, or 'stochastic' shocks from some assumed distribution used to perturb either the initial conditions or the conditioning variable. Adding random numbers drawn from the estimated in-sample error distributions to the scenario, but not the baseline, would be a further alternative. Different estimates of the uncertainty around each scenario trajectory will result from these, at most one of which could be correct.

The focus of our paper is on scenarios generated by deterministic shifts perturbing the marginal distributions of conditioning variables in estimated models. Typical examples would be examining an increase in interest rates in an economic model which took that policy variable as a given; or the impacts of climate change under a particular temperature scenario; or of 'social distancing' on the rate at which a pandemic might otherwise spread. However, more 'extreme' scenario changes could also be envisaged, such as considering different conditional models for the alternative scenario. Here possible examples would be a macroeconomic model still taking interest rates as the given policy instrument but contrasted with a model where interest rates are at the zero lower bound and quantitative easing (QE) is used instead as the policy change, involving a different transmission mechanism; or of complete lockdown on economic outcomes. Manifestly, the requirements for such scenarios to match the later outcome are considerably more demanding, albeit scenarios might be the only feasible approach to investigate their possible impacts in advance. Equally, a scenario outcome might generate a policy response, as seen in the UK when suggestions that half a million deaths might result from the then policy of trying to achieve 'herd immunity' to COVID-19 led to a new policy of 'social isolation', which changed the parameters of the spread process, leading to dramatically reduced predictions of deaths. Simultaneously, the new vastly higher predictions of unemployment induced a greatly increased financial boost from the government. Our paper analyzes the conditions needed for such scenarios to provide a reliable basis for policy.

Here we focus on inferences about the differences between scenario outcomes using projections based on models estimated in-sample. Let y_t be a policy-relevant variable determined by the conditional data generation process (DGP):

$$y_t = \beta_0 + \beta_1 x_{1,t} + \boldsymbol{\beta}_2' \mathbf{x}_{2,t} + \lambda y_{t-1} + \epsilon_t \ \text{ with } \ \epsilon_t \sim \mathsf{IN}\left[0, \sigma_\epsilon^2\right] \tag{1}$$

for $t = 1, \ldots, T$ where $|\lambda| < 1$ and the $\{x_{i,t}\}$ are valid conditioning variables, so that in-sample:

$$\mathsf{E}\left[y_t \mid x_{1,t}, \mathbf{x}_{2,t}, y_{t-1}\right] = \beta_0 + \beta_1 x_{1,t} + \boldsymbol{\beta}_2' \mathbf{x}_{2,t} + \lambda y_{t-1} \tag{2}$$

where $x_{1,t}$ is the variable to be altered by a scenario analysis and $\mathbf{x}_{2,t}$ are k conditioning variables with coefficients $\boldsymbol{\beta}_2$ also affecting y_t. We assume the scenario intervention takes the form of a shift of δ_x in the level of $x_{1,t}$ at time $T+1$. We denote the DGP outcome at $T+1$ in absence of this intervention by y_{T+1}, while the actual outcome following the intervention is labelled as y_{T+1}^*.

As the DGP is unknown, we consider estimating a model of y_t and subsequently creating projections with and without the scenario intervention δ_x. Using the estimated model, the predicted baseline in the absence of the intervention is denoted by $\widehat{y}_{T+1}$, while the predicted scenario projection is labelled $\widehat{y}_{T+1}^*$.

We are interested in conducting inference on the difference between the outcomes, $[y_{T+1}^* - y_{T+1}]$, using the estimated model and subsequently projected scenarios $[\widehat{y}_{T+1}^* - \widehat{y}_{T+1}]$. We consider two different general cases. First, when both the DGP and estimated model only include a single policy variable and no other covariates (beyond autoregressive dynamics), so $\boldsymbol{\beta}_2 = \mathbf{0}$ in (1). Second, we consider the case where additional covariates enter the DGP. We then study the properties of scenario projections using models that match the DGP, as well as models that are mis-specified for the DGP by incorrectly omitting relevant conditioning variables.

Answering the first question on whether testing the scenario difference yields additional insight beyond in-sample testing, we show that, when the model is an invariant linear relation in the policy variable, and the scenario intervention takes the form of a deterministic shift, then the statistical significance of the scenario difference is solely a function of the estimated coefficient on the policy variable in sample. The difference between scenarios will appear significant for any level of intervention δ_x if the estimated coefficient $\widehat{\beta}_1$ is itself deemed significantly different from zero. Testing on the scenario difference does not provide additional information beyond in-sample statistical testing of the estimated model.

Turning to the second question of when the estimated scenario differences yield unbiased estimates, in the more general case of multiple conditioning variables, the accuracy of scenario projections depends on the relationships between conditioning variables, and especially on their invariances to the intervention. In the absence of invariance, a well-specified model in-sample does not necessarily yield unbiased estimates of the scenario difference. Conversely, if there exists an invariant causal relationship between the policy variable and the additional conditioning variables, a mis-specified model omitting conditioning variables may be preferred unless the causal relationships between conditioning and policy variables are modelled. Crucially, invariance is not known *a-priori*, but can be tested in-sample as we discuss.

The structure of the paper is as follows. Section 2 considers scenario differences in a one-explanatory variable model to set the scene for testing differences between scenarios. Section 3 analyzes scenario differences under different model specifications and relationships between the policy and conditioning variables. Section 4 considers the applications of scenario analyses to wage share changes from unemployment, and agricultural impacts of climate change. Section 5 concludes. The appendix derives approximate variances of multi-period scenario outcomes. First, we consider an example to motivate our study.

1.1 A motivating example

Interval projections are regularly reported to encapsulate future uncertainties around a multi-period trajectory, as with the Bank of England's 'fan charts'. In such a setting, outcomes lying outside the interval projections, say 95% regions calculated from in-sample data, are taken to be significant deviations. Figure 1 illustrates this using artificial data where we know the DGP, which is the bivariate vector autoregression (VAR):

$$y_t = \beta_0 + \beta_1 x_{1,t} + \lambda y_{t-1} + \epsilon_t \tag{3}$$

$$x_{1,t} = \gamma + \nu_t \tag{4}$$

where $(\epsilon_t, \nu_t)' \sim \mathsf{IN}_2\left[\mathbf{0}, \boldsymbol{\Omega}\right]$, and $\boldsymbol{\Omega}$ is diagonal with $|\lambda| < 1$. For numerical calculations, the parameters $(\beta_0, \beta_1, \lambda, \gamma)$ are given the values$(0, 0.5, 0.3, 0)$, and $\omega_{11} = \omega_{22} = 1$ with $t = 1, \ldots, T$, where $T = 50$ defines the forecast origin, computing 10-steps ahead.

Figure 1 (panel a, top) records ten multi-step forecasts from the conditional model of the first variable, y_t (blue and dashed), matching the artificial DGP and taking the future values of $x_{1,T+1} \ldots x_{1,T+10}$ as known, with fan charts based on ± 1 and ± 2 forecast standard errors, denoted $\widehat{\sigma}_{f,h}$. The particular outcome for $y_{T+1} \ldots y_{T+10}$ drawn from that DGP is shown in black. In red and solid, we show a scenario trajectory for $y_{T+1} \ldots y_{T+10}$ where the variable x_1, controlled by an agency, is perturbed by the

magnitude $\delta_x = 2\omega_{22}$ from γ to $\gamma + \delta_x$, so is shifted up by that amount at each future observation relative to the baseline. One issue of concern is when can this new trajectory be deemed to be significantly different from the unperturbed outcomes? We will shortly consider the light grey line denoted 'DGP shifts with intervention'.
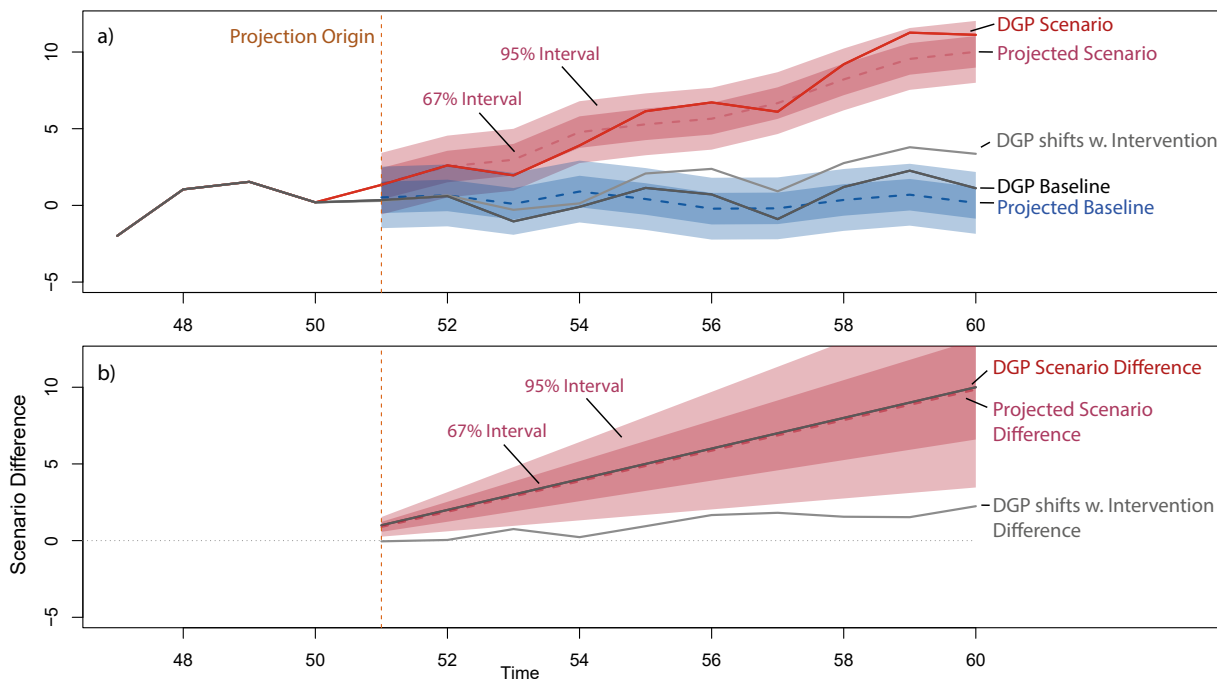

Figure 1: (a) Scenario projections with $\pm 2\widehat{\sigma}_{,fh}$ and $\pm 1\widehat{\sigma}_{f,h}$ for $h = 1, \ldots, 10$; and (b) the differences in the outcomes both under two states of nature.

Figure 1 (panel b) reports the differences between the baseline and scenario trajectories ten steps ahead (red). Because the comparison is between two trajectories based on the same in-sample estimated model, elements in common cancel, so the scenario difference is very smooth, always above zero, and is close to the difference the scenario perturbation would generate in the data (shown in dark gray).

However, Figure 1 also illustrates a setting where the scenario intervention unknowingly changes the parameters of the DGP (3) for $y_{T+1} \ldots y_{T+10}$ because of a failure of invariance to the intervention. This takes the form of reducing the forecast period β_1 by $0.2\omega_{11}$, which here is chosen to offset the agency's shift in $x_{1,T+1} \ldots x_{1,T+10}$, so the original $y_{T+1} \ldots y_{T+10}$, and perturbed outcomes $\widehat{y}^*_{T+1} \ldots \widehat{y}^*_{T+10}$, are essentially the same for all error draws. This offset could be interpreted as an extreme response of the 'private sector' to nullify a policy agency's intervention. The net effect is that, despite the projected difference in Figure 1(b), the scenario intervention does not lead to a different outcome, so the apparently significantly different trajectory is misleading. Consequently, had the agency implemented $\delta_x = 2\omega_{22}$ and forecast the future path as its scenario calculation in Figure 1(a), forecast failure would have resulted when the intervention was found to be ineffective. An important lesson from this analysis, emphasised below, is the key role of invariance of the model's parameters to the intervention if the outcome is to resemble the scenario. Fortunately, it may be feasible to test beforehand whether or not the model's parameters were altered on past occasions where earlier interventions were implemented (see e.g., Castle, Hendry, and Martinez, 2017). Discovering a lack of invariance could allow scenario calculations to be modified to more closely represent the outcome that will materialize.

2 Does testing the scenario difference provide insight beyond in-sample tests on an estimated univariate model?

First, we consider the simple case of a model matching the DGP in (3)–(4) where $\mathbf{\Omega}$ is diagonal but now need not be the unit matrix. Then assuming constant parameters in-sample:

$$\mathsf{E}\left[y_t\right] = \frac{\beta_0 + \beta_1\gamma}{1 - \lambda} = \mu$$

and using $\mathsf{V}\left[\cdot\right]$ to denote a variance:

$$\mathsf{V}\left[y_t\right] = \frac{\left(\omega_{11} + \beta_1^2\omega_{22}\right)}{1 - \lambda^2} = \sigma_{22}$$

with $\mathsf{E}\left[x_{1,t}\right] = \gamma$, $\mathsf{V}\left[x_{1,t}\right] = \omega_{22}$ and

$$\mathsf{E}\left[\left(x_{1,t} - \gamma\right)\left(y_{t-1} - \mathsf{E}\left[y_{t-1}\right]\right)\right] = 0 \tag{5}$$

The well-known limiting distribution of the estimates $\widehat{\beta}_1$ and $\widehat{\lambda}$ of the coefficients β_1 and λ is:

$$\begin{pmatrix} \sqrt{T}\left(\widehat{\beta}_1 - \beta_1\right) \\ \sqrt{T}\left(\widehat{\lambda} - \lambda\right) \end{pmatrix} \underset{\mathsf{D}}{\rightarrow} \mathsf{N}_2\left[\mathbf{0}, \sigma_\epsilon^2\mathbf{\Sigma}^{-1}\right] = \mathsf{N}_2\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \sigma_\epsilon^2\begin{pmatrix} \sigma_{11}^{-1} & 0 \\ 0 & \sigma_{22}^{-1} \end{pmatrix}\right] \tag{6}$$

where $\omega_{22} = \sigma_{11}$ here, and to highlight the key issues in scenario comparisons, we simplify by assuming that the sample size T is sufficiently large that $\widehat{\mathbf{\Sigma}} \approx \mathbf{\Sigma}$.

For given values of $x_{1,T+1}$ and y_T, the next period's baseline projection without a scenario perturbation is:

$$\widehat{y}_{T+1|T} = \widehat{\beta}_0 + \widehat{\beta}_1 x_{1,T+1} + \widehat{\lambda}y_T \tag{7}$$

so that:

$$\mathsf{E}\left[\widehat{y}_{T+1|T} \mid x_{T+1}, y_T\right] \approx \beta_0 + \beta_1 x_{T+1} + \lambda y_T.$$

After a scenario perturbation by δ_x, with parameters that are invariant to the intervention (a lack of invariance is considered in §3.3.3), the projected scenario outcome (denoted by *) is given by:

$$\widehat{y}_{T+1|T}^* = \widehat{\beta}_0 + \widehat{\beta}_1\left(x_{1,T+1} + \delta_x\right) + \widehat{\lambda}y_T \tag{8}$$

so that the estimated scenario is difference is given by:

$$\widehat{y}_{T+1|T}^* - \widehat{y}_{T+1|T} = \widehat{\beta}_1\delta_x \tag{9}$$

with:

$$\mathsf{E}\left[\widehat{y}_{T+1|T}^* - \widehat{y}_{T+1|T}\right] \approx \beta_1\delta_x$$

and:

$$\mathsf{V}\left[\widehat{y}_{T+1|T}^* - \widehat{y}_{T+1|T}\right] = \mathsf{V}\left[\widehat{\beta}_1\right]\delta_x^2 = \frac{\sigma_\epsilon^2\delta_x^2}{T\sigma_{11}}$$

as there is no correlation between the regressors. Then the t-statistic for the scenario difference is:

$$\mathsf{t}_{\beta_1\delta_x} = \frac{\widehat{\beta}_1\delta_x\sqrt{T\sigma_{11}}}{\sigma_\epsilon\delta_x} = \frac{\widehat{\beta}_1}{\mathsf{SE}\left[\widehat{\beta}_1\right]} = \mathsf{t}_{\widehat{\beta}_1} \tag{10}$$

which holds at all values of δ_x so does not depend on its magnitude. When $\widehat{\beta}_1$ is judged significantly different from zero (or not) in (10), all scenario changes at 1-step are significant (or not). This follows because the standard error of the difference $\widehat{y}^*_{T+1|T} - \widehat{y}_{T+1|T}$ is linear in δ_x, so the confidence interval around the scenario either never (or always) overlaps the origin for the chosen significance level (§2.2 considers stochastic scenario simulations).

Four other important points to note from this simple case are that:

(*i*) the stochastic error $\{\epsilon_t\}$ cancels in (9);

(*ii*) a vector of super-strong exogenous regressors $\mathbf{x}_{2,t}$ as in (1) (namely super exogenous as in Engle, Hendry, and Richard, 1983, and with no Granger causality from y_{t-i} or $x_{1,t-i}, i \geq 1$), would not affect (9), as such variables in common also cancel irrespective of whether or not they are included in the model;

(*iii*) the initial condition, y_T, cancels;

(*iv*) which still holds even if y_T is just an estimated initial condition $\widehat{y}_T$.

Thus, noise, omitted or included super-strong exogenous variables and mis-measured initial conditions do not affect the validity of the scenario analysis.

2.1 Scenario outcomes at $T + n$ for δ_x

The above results generalise to extending the previous case beyond a single time period for known future $\{x_{1,T+i}\}$, as the scenario projection n-periods out is:

$$\widehat{y}^*_{T+n|T} = \widehat{\beta}_0 \left(\sum_{i=0}^{n-1} \widehat{\lambda}^i \right) + \widehat{\beta}_1 \left(\sum_{i=1}^{n} \widehat{\lambda}^{n-i} x_{1,T+i} \right) + \widehat{\beta}_1 \left(\sum_{i=0}^{n-1} \widehat{\lambda}^i \right) \delta_x + \widehat{\lambda}^n y_T \tag{11}$$

whereas the baseline would be calculated as:

$$\widehat{y}_{T+n|T} = \widehat{\beta}_0 \left(\sum_{i=0}^{n-1} \widehat{\lambda}^i \right) + \widehat{\beta}_1 \left(\sum_{i=1}^{n} \widehat{\lambda}^{n-i} x_{1,T+i} \right) + \widehat{\lambda}^n y_T \tag{12}$$

which on the same assumptions as in the previous Section, leads to the scenario difference:

$$\widehat{y}^*_{T+n|T} - \widehat{y}_{T+n|T} = \widehat{\beta}_1 \left(\sum_{i=0}^{n-1} \widehat{\lambda}^i \right) \delta_x \tag{13}$$

since as before, all common elements cancel. The DGP scenario difference outcomes would be:

$$y^*_{T+n|T} - y_{T+n|T} = \beta_1 \left(\sum_{i=0}^{n-1} \lambda^i \right) \delta_x$$

so their variances would be zero. A scenario difference like (13) is a non-linear function of the in-sample parameter estimates, so the variance calculations are only approximate, and are derived in the Appendix for the model with a single policy variable in (3), showing that the intervention magnitude δ_x again cancels in a simple t-test of the scenario difference.

To simplify the formulae although the result holds more generally, we note the special case when $\omega_{11} = \omega_{22} = \sigma^2_\epsilon = 1$, $\beta_1 = 1$:

$$\mathsf{V}\left[\widehat{y}^*_{T+n|T} - \widehat{y}_{T+n|T} \right] \approx \frac{\left((1 - \lambda^n)^2 (1 - \lambda)^2 + \left(\left(1 - \left(n\lambda^{n-1} (1 - \lambda) + \lambda^n \right) \right)^2 \right) \right)}{T (1 - \lambda)^4} \delta^2_x \tag{14}$$

6

so for large n when $\lambda^n \approx n\lambda^{n-1} \approx 0$:

$$\mathsf{V}\left[\widehat{y}^*_{T+n|T} - \widehat{y}_{T+n|T}\right] \approx \frac{\delta_x^2}{T(1-\lambda)^4}\left(1 + (1-\lambda)^2\right) \tag{15}$$

Then combining (13) and (15), the t-statistic for the scenario difference is:

$$\mathsf{t}_{\beta_1\delta_x} = \frac{\widehat{\beta}_1\left(\sum_{i=0}^{n-1}\widehat{\lambda}^i\right)\delta_x}{\sqrt{\frac{\delta_x^2}{T(1-\lambda)^4}\left(1 + (1-\lambda)^2\right)}}$$

so again the significance will not depend on the magnitude of the perturbation δ_x as it cancels.

In the practical setting where stochastic $\{x_{1,T+i}\}$ will not be known into the future, a model thereof will need to be developed for multi-period forecasting, or direct forecasting must be used (see Chevillon and Hendry, 2005), where there will be a considerable increase in uncertainty the larger n is.

2.2 Stochastic scenario intervention

The above analysis focused on a deterministic perturbation δ_x. If the scenario intervention δ_x is itself stochastic, the mean scenario difference could be estimated using $\bar{\delta}_x = L^{-1}\sum_{i=1}^{L}\delta_{x,i}$ where L is the number of scenario draws (e.g., draws across temperature outcomes in a climate-impacts model as in Pretis et al. 2018) with the resulting average scenario difference given by $\widehat{\beta}_1\bar{\delta}_x$. When $\delta_{i,x} \sim \mathsf{IID}[\delta_x, \sigma_\delta^2]$ and $\bar{\delta}_x$ and $\widehat{\beta}_1$ are independent, then the variance of the estimated scenario difference can be approximated by $\mathsf{V}[\widehat{\beta}_1\bar{\delta}_x] \approx \mathsf{E}[\bar{\delta}_x^2]\mathsf{V}[\widehat{\beta}_1] + \mathsf{E}[\widehat{\beta}_1^2]\mathsf{V}[\bar{\delta}_x]$. In the absence of non-linear effects of $x_{1,t}$, or non-zero covariances between δ_x and β_1, a joint test, or joint sampling over δ_x and $\widehat{\beta}_1$, to assess the significance of scenario deviations will not be more informative than a single test on β_1. If $\beta_1 \neq 0$ then for any value of δ_x the scenario difference will be non-zero. Thus, sampling over both δ_x and β_1, or δ_y and λ, will not add information beyond testing on β_1 or λ. However, using independent stochastic perturbations with means of δ_x and δ_y will increase the variances around $\widehat{y}^*_{T+1|T}$ and $\widetilde{y}^*_{T+1|T}$ (corresponding to perturbing the dependent variable directly), so scenario differences could be insignificant even when $\mathsf{t}_{\widehat{\beta}_1}$ or $\mathsf{t}_{\widehat{\lambda}}$ rejected their nulls.

2.3 Scenario outcomes at $T+n$ for δ_y

Next, we consider a scenario intervention perturbing the forecast-origin value of the dependent variable y_T by δ_y and assess the variance for the n-step ahead projection. For a single step ahead, consider retaining (7) but changing (8) to:

$$\widetilde{y}^*_{T+1|T} = \widehat{\beta}_0 + \widehat{\beta}_1 x_{1,T+1} + \widehat{\lambda}(y_T + \delta_y) \tag{16}$$

so that:

$$\widetilde{y}^*_{T+1|T} - \widehat{y}_{T+1|T} = \widehat{\lambda}\delta_y \tag{17}$$

then a similar analysis results at $T+1$ replacing $\widehat{\beta}_1\delta_x$ by $\widehat{\lambda}\delta_y$.

The multi-period baseline (12) is unchanged but the perturbed scenario (11) becomes:

$$\widehat{y}^*_{T+n|T} = \widehat{\beta}_0\left(\sum_{i=0}^{n-1}\widehat{\lambda}^i\right) + \widehat{\beta}_1\left(\sum_{i=1}^{n}\widehat{\lambda}^{n-i}x_{T+i}\right) + \left(\sum_{i=1}^{n}\widehat{\lambda}^i\right)\delta_y + \widehat{\lambda}^n y_T \tag{18}$$

7

so that:

$$\widehat{y}^*_{T+n|T} - \widehat{y}_{T+n|T} = \left(\sum_{i=1}^{n} \widehat{\lambda}^i\right)\delta_y \tag{19}$$

Using the same approximations for multi-step variances as above, but now only depending on $\widehat{\lambda}$:

$$\mathsf{V}\left[\widehat{y}^*_{T+n|T} - \widehat{y}_{T+n|T}\right] \approx \frac{\sigma^2_\epsilon \lambda^2 \left(1 - \lambda^n\right)^2 \delta^2_y}{T\sigma_{11}\left(1 - \rho^2\right)\left(1 - \lambda\right)^2} \tag{20}$$

In the special case $\sigma_{11} = \sigma_{22} = 1$, $\beta_1 = 1$, $\sigma_{12} = \rho = 0$:

$$\mathsf{V}\left[\widehat{y}^*_{T+n|T} - \widehat{y}_{T+n|T}\right] \approx \frac{\lambda^2 \left(1 - \lambda^n\right)^2 \delta^2_y}{T\left(1 - \lambda\right)^2} \tag{21}$$

As above, the intervention δ_y cancels when testing the scenario difference. For $\lambda = 0.85$ and $n = 8$, (21) equals $0.17\delta^2_y \approx 11.0$ for $\delta^2_y = 9\sigma^2_y$ at $T = 100$ whereas (14) equals $2.56\delta^2_x \approx 23.0$, even though at $\beta_1 = 1$, (13) and (19) only differ by a factor of $\widehat{\lambda}$. Thus, the degree of persistence mediates the multi-step outcomes both from the cumulative impact of a perturbation, and whether y_T or x_{T+1} is changed.

2.4 Recommendations

In summary, for the single regressor dynamic model, testing the scenario difference does not yield additional insight beyond testing the significance of the estimated in-sample parameters when the model is linear in the perturbed known future scenario variable. This result holds for both single and multiple periods ahead, and for perturbations in both the lagged dependent as well as independent variable. We now consider scenario differences in processes with additional covariates and analyze the four possible ways in which the invariance to the intervention of relationships between covariates might fail.

3 When do estimated scenario differences yield unbiased estimates of the true differences in outcomes?

We now consider the case where the DGP includes both the policy variable perturbed by an intervention as well as k additional covariates $\mathbf{x}_2$ as in (1). Such interventions would correspond to counter-factual experiments when the agency controls $x_{1,t}$ to influence y_t, but can be more general where an agency wishes to explore possible futures. In this section, we consider the cases where the estimated model coincides with the DGP (1) as well as where the model is mis-specified for the DGP by omitting conditioning variables. A crucial factor affecting the differences in scenario outcomes is the relationship between the policy variable $x_{1,t}$ and the covariates $\mathbf{x}_2$, even when the $\mathbf{x}_{2,T}$ are independent of the DGP error on the equation of interest.

In the absence of any scenario intervention at time $T + 1$ the DGP outcome from (1) would be:

$$y_{T+1} = \beta_0 + \beta_1 x_{1,T+1} + \boldsymbol{\beta}'_2 \mathbf{x}_{2,T+1} + \lambda y_T + \epsilon_{T+1} \tag{22}$$

The scenario intervention is again one where $x_{1,T+1}$ is perturbed by δ_x for invariant parameters in (22), leading to:

$$y^*_{T+1} = \beta_0 + \beta_1 \left(x_{1,T+1} + \delta_x\right) + \boldsymbol{\beta}'_2 \mathbf{x}^*_{2,T+1} + \lambda y_T + \epsilon_{T+1} \tag{23}$$

where $\mathbf{x}^*_{2,T+1}$ reflects any impact of δ_x on $\mathbf{x}_{2,T+1}$ through its links to $x_{1,T+1}$. The linkage between $\mathbf{x}_2$ and x_1 is described by the projection:

$$\mathbf{x}_{2,t} = \boldsymbol{\pi}_0 + \boldsymbol{\pi}_1 x_{1,t} + \boldsymbol{\pi}_2 y_{t-1} + \mathbf{u}_t \quad \text{with} \quad \mathbf{u}_t \sim \mathsf{IN}_k\left[\mathbf{0}, \sigma^2_u \boldsymbol{\Omega}\right] \tag{24}$$

which may, but need not, be the DGP for $\{\mathbf{x}_{2,t}\}$ (in practice, $\mathbf{x}_{2,t-1}$ is likely to be relevant). Then, comparing (22) with (23), the scenario difference in the DGP is:

$$y^*_{T+1} - y_{T+1} = \beta_1 \delta_x + \boldsymbol{\beta}'_2 \left(\mathbf{x}^*_{2,T+1} - \mathbf{x}_{2,T+1} \right) \tag{25}$$

where $\mathbf{x}^*_{2,T+1} = \mathbf{x}_{2,T+1}$ reproduces the earlier result that other super-strongly exogenous regressors cancel. Even if the conditional DGP (22) remains invariant to the scenario intervention, the outcome depends on the reaction of $\mathbf{x}_{2,T+1}$ to an intervention on $x_{1,T+1}$. For example, the Bank of England uses interest rates (x_1) to change inflation (y), and that operates though the impact of interest rates on aggregate demand ($\mathbf{x}_2$), which in turn affects inflation, and as (25) shows, does not require $\beta_1 \neq 0$. As before, the direct impacts from y_T and the error ϵ_{T+1} from (22) with (23) cancel between the scenarios, so do not affect (25) for 1-step ahead, but that will change if any DGP parameters are not invariant to the intervention, and at multi-steps ahead, especially when there are feedbacks from y_{t-1} onto any of the $x_{i,t}$.

A possible situation is one of co-breaking between $\mathbf{x}_{2,t}$ and $x_{1,t}$ such that $\boldsymbol{\beta}'_2 \boldsymbol{\pi}_1 = \mathbf{0}$: see Hendry and Massmann (2007). As:

$$\mathbf{x}^*_{2,T+1} = \boldsymbol{\pi}_0 + \boldsymbol{\pi}_1 \left(x_{1,T+1} + \delta_x \right) + \boldsymbol{\pi}_2 y_T + \mathbf{u}_{T+1} = \mathbf{x}_{2,T+1} + \boldsymbol{\pi}_1 \delta_x$$

then:

$$y^*_{T+1} - y_{T+1} = \left(\beta_1 + \boldsymbol{\beta}'_2 \boldsymbol{\pi}_1 \right) \delta_x = \beta_1 \delta_x \tag{26}$$

leading to the same outcome for $y^*_{T+1} - y_{T+1}$ in (25) as if $\mathbf{x}^*_{2,T+1} = \mathbf{x}_{2,T+1}$ even though $\mathbf{x}^*_{2,T+1}$ has shifted. Partial co-breaking would lead to some cancellation. A test of in-sample co-breaking is to add any impulse and step indicators for outliers and shifts in $x_{1,t}$ detected in estimates of (24) to (1) and check their significance: if $\boldsymbol{\beta}'_2 \boldsymbol{\pi}_1 = \mathbf{0}$, they should be irrelevant. However, unless k is very small, it seems unlikely that all regressors would co-break precisely with shifts in $x_{1,t}$ given their different parameter values.

3.1 Model matching the DGP

When the model matches the DGP, in an optimistic scenario where the $\mathbf{x}_{2,T+1}$ are known, the projected values are given by $\widehat{y}_{T+1}$ and $\widehat{y}^*_{T+1}$ for the unperturbed and perturbed scenario respectively, so the estimated scenario difference is the same as in Section 2.

3.2 Mis-specified model

A more likely setting is for the model to be mis-specified for the DGP, crucially by omitting relevant variables. We focus on the omitted variable case where we let the mis-specified model omit all $\mathbf{x}_{2,t}$:

$$y_t = \phi_0 + \phi_1 x_{1,t} + \psi y_{t-1} + e_t \text{ assuming } e_t \sim \mathsf{IN} \left[0, \sigma_e^2 \right] \tag{27}$$

with $|\psi| < 1$. Using (1) and assuming the link between variables in (24) holds in-sample, the mapping between the DGP parameters and the model's coefficients is:

$$y_t = \left(\beta_0 + \boldsymbol{\beta}'_2 \boldsymbol{\pi}_0 \right) + \left(\beta_1 + \boldsymbol{\beta}'_2 \boldsymbol{\pi}_1 \right) x_{1,t} + \left(\boldsymbol{\beta}'_2 \boldsymbol{\pi}_2 + \lambda \right) y_{t-1} + \left(\boldsymbol{\beta}'_2 \mathbf{u}_t + \epsilon_t \right) \tag{28}$$

so that:

$$\phi_0 = \left(\beta_0 + \boldsymbol{\beta}'_2 \boldsymbol{\pi}_0 \right); \quad \phi_1 = \left(\beta_1 + \boldsymbol{\beta}'_2 \boldsymbol{\pi}_1 \right); \quad \text{and} \quad \psi = \left(\boldsymbol{\beta}'_2 \boldsymbol{\pi}_2 + \lambda \right).$$

These correspond to the standard results for omitted variable biases in linear models. From a sample of $t = 1, \ldots, T$, and having taken deviations from sample means, an investigator again estimates the coefficients ϕ_1 and ψ as $\widehat{\phi}_1$ and $\widehat{\psi}$ where for large T:

$$
\begin{pmatrix} \sqrt{T} \left(\widehat{\phi}_1 - \phi_1 \right) \\ \sqrt{T} \left(\widehat{\psi} - \psi \right) \end{pmatrix} \sim \mathsf{N}_2 \left[\mathbf{0}, \sigma_e^2 \mathbf{\Sigma}^{-1} \right] \tag{29}
$$

and we assume $\widehat{\mathbf{\Sigma}} \approx \mathbf{\Sigma}$. Subsequently, the estimated mis-specified model is used to create projections, where the scenario projected values of the mis-specified model are denoted by $\widetilde{y}_{T+1}^{*}$ and $\widetilde{y}_{T+1}$ with and without perturbation.

3.2.1 Mis-specified projection in the absence of intervention

For given values of $x_{1,T+1}$ and y_T with invariant parameters, the baseline projection for the next period using the mis-specified model is:

$$
\widetilde{y}_{T+1|T} = \widehat{\phi}_0 + \widehat{\phi}_1 x_{1,T+1} + \widehat{\psi} y_T \tag{30}
$$

We assume that T is sufficiently large that:

$$
\mathsf{E} \left[\widetilde{y}_{T+1|T} \mid x_{1,T+1}, \mathbf{x}_{2,T+1}, y_T \right] \approx \phi_0 + \phi_1 x_{1,T+1} + \psi y_T
$$

because finite-sample biases will be small compared to the effects of mis-specification and shifts. Using (30) in the absence of a scenario intervention, from (28) the mis-specification of the model (27) for the DGP (1) leads to a mean prediction error in the level relative to the DGP outcome:

$$
\begin{aligned}
\mathsf{E} \left[\left(y_{T+1} - \widetilde{y}_{T+1|T} \right) \mid x_{1,T+1}, \mathbf{x}_{2,T+1}, y_T \right] &= (\beta_0 - \phi_0) + (\beta_1 - \phi_1) x_{1,T+1} + \boldsymbol{\beta}_2' \mathbf{x}_{2,T+1} + (\lambda - \psi) y_T \\
&= \boldsymbol{\beta}_2' \left(\mathbf{x}_{2,T+1} - \boldsymbol{\pi}_0 - \boldsymbol{\pi}_1 x_{1,T+1} - \boldsymbol{\pi}_2 y_T \right) = \boldsymbol{\beta}_2' \mathbf{u}_{T+1}
\end{aligned}
$$

which will be distributed around zero. As shown in Clements and Hendry (1998), in a constant-parameter world, omitted variables do not induce serious problems, here merely augmenting the usual forecast-error variance with an additional variance of $\sigma_u^2 \boldsymbol{\beta}_2' \mathbf{\Omega} \boldsymbol{\beta}_2$.

3.2.2 Mis-Specified projection with an intervention

Undertaking a scenario study in the mis-specified model where $x_{1,T+1}$ is perturbed by δ_x leads to:

$$
\widetilde{y}_{T+1|T}^{*} = \widehat{\phi}_0 + \widehat{\phi}_1 \left(x_{1,T+1} + \delta_x \right) + \widehat{\psi} y_T \tag{31}
$$

The impact from changing $x_{1,T+1}$ by δ_x entails calculating the estimated scenario difference as $\widetilde{y}_{T+1|T}^{*} - \widetilde{y}_{T+1|T} = \widehat{\phi}_1 \delta_x$, with an expected value $\mathsf{E}[\widehat{\phi}_1 \delta_x] \approx \phi_1 \delta_x = \left(\beta_1 + \boldsymbol{\beta}_2' \boldsymbol{\pi}_1 \right) \delta_x$. Comparing the baseline (30) with the scenario intervention (31) yields the expected difference:

$$
\begin{aligned}
\mathsf{E} \left[\left(\widetilde{y}_{T+1|T}^{*} - \widetilde{y}_{T+1|T} \right) \mid x_{1,T+1} \right] &= \phi_0 + \phi_1 \left(x_{1,T+1} + \delta_x \right) + \lambda y_T - \phi_0 - \phi_1 x_{1,T+1} - \psi y_T \\
&= \left(\beta_1 + \boldsymbol{\beta}_2' \boldsymbol{\pi}_1 \right) \delta_x
\end{aligned} \tag{32}
$$

Importantly, this is a unique counter-factual prediction of the 1-step ahead difference between the scenarios when $\mathbf{x}_{2,T+1}$ and the link (24) between $\mathbf{x}_2$ and x_1 (or generalizations thereof) are unknown to the

agency but nevertheless invariant to the intervention. Then, the variance $\mathsf{V}[\cdot]$ of the scenario difference $\widetilde{y}^*_{T+1|T} - \widetilde{y}_{T+1|T} = \widehat{\phi}_1 \delta_x$ between (31) and (30) is estimated under the null of correct specification as:

$$\mathsf{V}\left[\widetilde{y}_{T+1|T} - \widehat{y}_{T+1|T}\right] = \delta_x^2 \mathsf{V}\left[\widehat{\phi}_1\right] = \frac{\sigma_\epsilon^2 \delta_x^2}{T\sigma_{11}\left(1 - \psi^2\right)} \tag{33}$$

While the investigator may anticipate a scenario difference of $\widehat{\phi}_1 \delta_x$ with the variance in (33), the actual outcome will differ depending on the actual links between covariates $x_{1,T+1}$ and $\mathbf{x}_{2,T+1}$ as we now consider.

3.3 Linkages between covariates and their impacts on scenario projections

We now investigate four possible forms of linkage between $x_{1,T+1}$ and $\mathbf{x}_{2,T+1}$ when creating scenario projections using a model that matches the DGP, as well as the above model being mis-specified for the DGP. First, we consider the case where $\mathbf{x}_{2,T+1}$ is unaffected by an intervention in $x_{1,T+1}$. Second, the case when there exists an invariant causal link between $\mathbf{x}_{2,T+1}$ and $x_{1,T+1}$. Third, there exists an in-sample relationship between $\mathbf{x}_{2,T+1}$ and $x_{1,T+1}$, however, this relationship is not invariant to interventions. Fourth, the relationship between y_t and $x_{1,t}$ itself is not invariant to interventions. The analysis here just considers scenario projections for one-period ahead, as for two-periods ahead and beyond, most formulae cease to yield useful insights.

3.3.1 Covariates $\mathbf{x}_2$ unaffected by the intervention on x_1

DGP outcome: When all $\mathbf{x}_{2,T+1}$ are unaffected by the intervention, so that $\mathbf{x}^*_{2,T+1} = \mathbf{x}_{2,T+1}$, then from (25) the scenario difference in the DGP is identical to the univariate case in section 2:

$$y^*_{T+1} - y_{T+1} = \beta_1 \delta_x \tag{34}$$

This is often the assumption under which scenarios are calculated, basically assuming *ceteris paribus*.

Model matches the DGP: Again when $\mathbf{x}^*_{2,T+1} = \mathbf{x}_{2,T+1}$ in (25), even if (24) is merely an in-sample projection, then the actual effect will also be $\beta_1 \delta_x$, which is estimated without bias using the model matching the DGP.

Model does not match the DGP: Now the scenario prediction using the mis-specified model will be incorrect on average by $\boldsymbol{\beta}'_2 \boldsymbol{\pi}_1 \delta_x$. This could have either sign and potentially any magnitude. This directly corresponds to the well-known result of omitted variable bias in linear models.

3.3.2 Invariant causal relation between x_1 and $\mathbf{x}_2$

DGP outcome: When the in-sample linkage between x_1 and $\mathbf{x}_2$ is the invariant causal relation given by (24) then $\mathbf{x}^*_{2,T+1} = \mathbf{x}_{2,T+1} + \boldsymbol{\pi}_1 \delta_x$, and the correct outcome (25) becomes:

$$y^*_{T+1} - y_{T+1} = \left(\beta_1 + \boldsymbol{\beta}'_2 \boldsymbol{\pi}_1\right) \delta_x \tag{35}$$

As noted above, co-breaking by $\boldsymbol{\beta}_2$ may lead to some or all elements of $\boldsymbol{\beta}'_2 \boldsymbol{\pi}_1$ being zero.

11

Model matches the DGP: In the absence of correctly modelling the causal relation between x_1 and $\mathbf{x}_{2,t}$ the model matching the pre-intervention DGP results in a biased estimate of the scenario difference:

$$\mathsf{E}\left[\left(\widehat{y}_{T+1}^* - \widehat{y}_{T+1}\right) - \left(y_{T+1}^* - y_{T+1}\right) \mid x_{1,T+1}\right] = (\beta_1 \delta_x) - (\beta_1 + \boldsymbol{\beta}_2' \boldsymbol{\pi}_1)\delta_x = \boldsymbol{\beta}_2' \boldsymbol{\pi}_1 \delta_x \qquad (36)$$

This is the same as the error on the outcome in §3.3.1 when the model does **not** match the DGP. To correctly calculate the future outcomes would require taking account of the relationship in (24), via a system dynamic simulation. Replacing $x_{1,t}$ in (24) by $x_{1,t-1}$ would just alter the timing at which (35) affected the outcome.

Model does not match the DGP: The impact from changing $x_{1,T+1}$ by δ_x entails calculating the scenario difference as $\widetilde{y}_{T+1|T}^* - \widetilde{y}_{T+1|T} = \widehat{\phi}_1 \delta_x$, with an expected value $\mathsf{E}[\widehat{\phi}_1 \delta_x] \approx \phi_1 \delta_x = \left(\beta_1 + \boldsymbol{\beta}_2' \boldsymbol{\pi}_1\right) \delta_x$, as comparing (30) with (31):

$$\mathsf{E}\left[\left(\widetilde{y}_{T+1|T}^* - \widetilde{y}_{T+1|T}\right) \mid x_{1,T+1}\right] = \phi_0 + \phi_1 \left(x_{1,T+1} + \delta_x\right) + \lambda y_T - \phi_0 - \phi_1 x_{1,T+1} - \psi y_T$$
$$= \left(\beta_1 + \boldsymbol{\beta}_2' \boldsymbol{\pi}_1\right) \delta_x \qquad (37)$$

which matches the DGP outcome in (35). When (24) is an invariant causal relation, so $x_{1,T+1}$ and $\mathbf{x}_{2,T+1}$ co-break in the constant relation (24) with $\mathbf{x}_{2,T+1}^* = \mathbf{x}_{2,T+1} + \boldsymbol{\pi}_1 \delta_x$, and maintaining the assumption of parameter invariance in (1), then the difference between the projected scenario and the true scenario outcome is on average:

$$\mathsf{E}\left[\left(y_{T+1}^* - \widetilde{y}_{T+1|T}^*\right) \mid x_{1,T+1}, \mathbf{x}_{2,T+1}, y_T\right] = \beta_0 + \beta_1 \left(x_{1,T+1} + \delta_x\right) + \boldsymbol{\beta}_2' \mathbf{x}_{2,T+1}^* + \lambda y_T$$
$$- \phi_0 - \phi_1 \left(x_{1,T+1} + \delta_x\right) - \psi y_T$$
$$= \boldsymbol{\beta}_2' \mathbf{u}_{T+1}$$

Consequently, despite the mis-specification, given invariant parameters in the rest of the system DGP under the scenario change, and a constant causal link between the included and unknowingly excluded variables in both states of nature, the **correct scenario calculation** of $(\beta_1 + \boldsymbol{\beta}_2' \boldsymbol{\pi}_1)\delta_x$ in (37) results. This finding that a mis-specified model leads to an unbiased scenario prediction is closely related to the concept of conditioning on a post-treatment variable in the causal-inference literature, and to deliberately omitting exogenous variables from an open forecasting model as explained in Hendry and Mizon (2012). As the outcome of interest is the effect of x_1 onto y_t, if $\mathbf{x}_{2,t}$ itself is affected by x_1, then $\mathbf{x}_{2,t}$ should not be included in a projection model for y_t unless the relationship between x_1 and $\mathbf{x}_{2,t}$ is also formally modelled. While causality is difficult to establish in empirical models, invariance can be tested using the approach discussed in §3.5 and applied below.

3.3.3   In-sample projection and no invariant causal relation between x_1 and $\mathbf{x}_2$

The requirements of an invariant causal relationship are very strong and most unlikely to hold in a wide-sense non-stationary world facing intermittent shifts of distributions. When the relationship between x_1 and $\mathbf{x}_{2,t}$ in (24) holds in-sample merely as a projection based on inter-correlations, rather than as an invariant relation, then in the absence of any intervention, (24) would produce:

$$\mathbf{x}_{2,T+1} = \boldsymbol{\pi}_0 + \boldsymbol{\pi}_1 x_{1,T+1} + \boldsymbol{\pi}_2 y_T + \mathbf{u}_{T+1}$$

but this link is in fact not invariant to the intervention on $x_{1,T+1}$ of δ_x, so alters the parameters of (24) to:

$$\mathbf{x}_{2,T+1}^* = \boldsymbol{\pi}_0^* + \boldsymbol{\pi}_1^* \left(x_{1,T+1} + \delta_x\right) + \boldsymbol{\pi}_2^* y_T + \mathbf{u}_{T+1} \qquad (38)$$

Then using (38):

$$\mathbf{x}_{2,T+1}^* - \mathbf{x}_{2,T+1} = \boldsymbol{\pi}_1 \delta_x + (\boldsymbol{\pi}_0^* - \boldsymbol{\pi}_0) + (\boldsymbol{\pi}_1^* - \boldsymbol{\pi}_1)(x_{1,T+1} + \delta_x) + (\boldsymbol{\pi}_2^* - \boldsymbol{\pi}_2) y_T \qquad (39)$$

Consequently, from (25):

$$\begin{aligned}
y_{T+1}^* - y_{T+1} = {} & \left(\beta_1 + \boldsymbol{\beta}_2' \boldsymbol{\pi}_1\right) \delta_x + \boldsymbol{\beta}_2' \left(\boldsymbol{\pi}_0^* - \boldsymbol{\pi}_0\right) \\
& + \boldsymbol{\beta}_2' \left(\boldsymbol{\pi}_1^* - \boldsymbol{\pi}_1\right) (x_{1,T+1} + \delta_x) + \boldsymbol{\beta}_2' \left(\boldsymbol{\pi}_2^* - \boldsymbol{\pi}_2\right) y_T
\end{aligned} \qquad (40)$$

so the scenario difference now depends on the initial condition as well as on all the parameter shifts and could be badly mis-estimated by $\beta_1 \delta_x$.

Model does not match the DGP: When the relation in (24) is merely an in-sample projection, and instead (38) occurs, either because of the shift in δ_x or from other changes, then a major difference emerges between the responses $(\beta_1 + \boldsymbol{\beta}_2' \boldsymbol{\pi}_1) \delta_x$ predicted by the model and that from the DGP as $(y_{T+1}^* - y_{T+1})$ in (40) leads to $(\beta_1 + \boldsymbol{\beta}_2' \boldsymbol{\pi}_1) \delta_x + \boldsymbol{\beta}_2' (\boldsymbol{\pi}_0^* - \boldsymbol{\pi}_0) + \boldsymbol{\beta}_2' (\boldsymbol{\pi}_1^* - \boldsymbol{\pi}_1)(x_{1,T+1} + \delta_x)$. Thus:

$$\mathsf{E}\left[\left(y_{T+1}^* - \widetilde{y}_{T+1|T}^* \right) \mid x_{1,T+1}, \mathbf{x}_{2,T+1}, y_T \right] = \boldsymbol{\beta}_2' \left(\boldsymbol{\pi}_0^* - \boldsymbol{\pi}_0 \right) + \boldsymbol{\beta}_2' \left(\boldsymbol{\pi}_1^* - \boldsymbol{\pi}_1 \right) (x_{1,T+1} + \delta_x) \qquad (41)$$

so the calculated and actual scenario responses could be wildly different. Consequently, in the absence of an invariant causal relationship, both the well-specified and mis-specified models yield unreliable estimates for the scenario difference.

3.3.4 Scenario intervention shifts the DGP itself

The situation where changes in $x_{1,t}$ not only shift $\mathbf{x}_{2,t}$ and also alter the DGP parameters in (1) is more complicated still. In place of (23), to illustrate, we just consider the special case:

$$y_{T+1}^* = \beta_0^* + \beta_1^* (x_{1,T+1} + \delta_x) + \boldsymbol{\beta}_2' \mathbf{x}_{2,T+1}^* + \lambda y_T + \epsilon_{T+1} \qquad (42)$$

where δ_x not only changes the parameters in the model equation for y but also all of the parameters in the relationship with $\mathbf{x}_2$ as in (38) so that:

$$\begin{aligned}
y_{T+1}^* - y_{T+1} = {} & (\beta_0^* - \beta_0) + \beta_1 \delta_x + (\beta_1^* - \beta_1)(x_{1,T+1} + \delta_x) + \boldsymbol{\beta}_2' \left(\mathbf{x}_{2,T+1}^* - \mathbf{x}_{2,T+1} \right) \\
= {} & \left(\beta_1 + \boldsymbol{\beta}_2' \boldsymbol{\pi}_1 \right) \delta_x + \boldsymbol{\beta}_2' \left(\boldsymbol{\pi}_0^* - \boldsymbol{\pi}_0 \right) + (\beta_0^* - \beta_0) \\
& + \left(\beta_1^* - \beta_1 + \boldsymbol{\beta}_2' \left(\boldsymbol{\pi}_1^* - \boldsymbol{\pi}_1 \right) \right) (x_{1,T+1} + \delta_x) + \boldsymbol{\beta}_2' \left(\boldsymbol{\pi}_2^* - \boldsymbol{\pi}_2 \right) y_T
\end{aligned} \qquad (43)$$

Other parameter changes would make (43) more complicated still, but the principle that all failures of invariance add yet more terms is clear. Indeed, even if the linkage in (24) was an invariant causal relation, (43) would become:

$$y_{T+1}^* - y_{T+1} = \left(\beta_1 + \boldsymbol{\beta}_2' \boldsymbol{\pi}_1 \right) \delta_x + (\beta_0^* - \beta_0) + (\beta_1^* - \beta_1)(x_{1,T+1} + \delta_x) \qquad (44)$$

still leading to a complicated outcome very different from might be anticipated.

Model does not match the DGP: The problem seen in (41) is of course magnified by any shifts in the DGP parameters themselves. The calculated scenario remains $\left(\beta_1 + \boldsymbol{\beta}_2' \boldsymbol{\pi}_1 \right) \delta_x$ whereas the DGP difference will be (43), so the two need not be alike in any way.

3.4 Recommendations

Overall, even if the in-sample conditional DGP equation for the target variable is known, the outcomes of scenario interventions will depend on the links between the perturbed policy variable and any other relevant variables, as well as the dynamics, and will reflect all failures of invariance in both the DGP equation and the DGPs for the regressor variables.

Table 1 summarises the main cases, where **#** simply denotes too complicated to be recorded in detail, but differing across the columns. In short, if conditioning variables are unaffected by the policy variable, then a model matching the DGP is preferred. However, if there is an invariant causal link between the variables, then a mis-specified model omitting all other regressors can yield unbiased estimates of the scenario difference, while a model matching the DGP may not. Noting that scenario differences essentially correspond to total effects, whereas just perturbing the policy variable and not including its impacts on other regressors just delivers a partial effect, clarifies what might otherwise seem an odd result. When some but not all of the other relevant variables are unknowingly omitted, biased estimates of the scenario difference will usually result.

Table 1: Actual and Predicted Difference between perturbed and unperturbed outcomes.

Case	DGP Scenario Difference	Model=DGP	Mis-specified Model
1: $\mathbf{x}_2$ unaffected	$\beta_1 \delta_x$	$\widehat{\beta_1} \delta_x$	$\widetilde{(\beta_1 + \boldsymbol{\beta}_2' \boldsymbol{\pi}_1)} \delta_x$
2: $x_1, \mathbf{x}_2$ invariant	$\left(\beta_1 + \boldsymbol{\beta}_2' \boldsymbol{\pi}_1\right) \delta_x$	$\widetilde{\boldsymbol{\beta}_2' \boldsymbol{\pi}_1} \delta_x$	$\widetilde{(\beta_1 + \boldsymbol{\beta}_2' \boldsymbol{\pi}_1)} \delta_x$
3: $x_1, \mathbf{x}_2$ not invariant	#	#	#
4: DGP not invariant	#	#	#

As invariance properties are generally unknown from the outset, a key lesson from this analysis is the crucial requirement to model all the in-sample linkages and test for any previous shifts in such relationships. The challenge of assessing invariance is not insurmountable, as considered by automatic testing of interventions in marginal models proposed in Hendry and Santos (2010) and Castle, Hendry, and Martinez (2017). While that will not be possible for any variables in $\mathbf{x}_{2,t}$ that are either deliberately or inadvertently omitted from the model, formally testing for parameter invariance remains viable unless such omitted variables are both orthogonal to all included ones and have zero population means.

3.5 Testing for parameter invariance in the marginal models

A feature of these analyses is the need for models to be invariant to the intervention if the scenario is to match reality. The marginal variable above is $x_{1,t}$ and although building a complete model thereof is usually infeasible, the most deleterious failures of invariance concern induced location shifts captured in (43) by $\boldsymbol{\beta}_2' (\boldsymbol{\pi}_0^* - \boldsymbol{\pi}_0)$. Thus, consider detecting interventions (location shifts or impulses) in a marginal model of $x_{1,t}$ as in (45) where m could be 0 or 1 depending on the exogeneity status of $\mathbf{x}_{2,t}$:

$$x_{1,t} = \gamma_0 + \sum_{j=0}^{s} \left(\gamma_{1,j-1} x_{1,t-j-1} + \gamma_{2,j} y_{t-j-1} + \boldsymbol{\Gamma}_{2,j-1} \mathbf{x}_{2,t-j-m} \right) + \sum_{i=1}^{T} \tau_i 1_{\{i \leq t\}} + u_t \qquad (45)$$

where $1_{\{i \leq t\}}$ is a step-indicator equal to unity till time t then zero thereafter. Selection at a chosen level of significance p_α reduces the set of T indicators to a subset of m detected shifts at times t_i. Subsequently these detected shifts are entered in conditional models of $\mathbf{x}_{2,t}$ in (46):

$$\mathbf{x}_{2,t} = \boldsymbol{\pi}_0 + \sum_{j=0}^{s} \left(\boldsymbol{\pi}_{1,j} x_{1,t-j} + \boldsymbol{\pi}_{2,j} y_{t-j-1} + \boldsymbol{\Pi}_{2,j} \mathbf{x}_{2,t-j-1} \right) + \sum_{i=1}^{m} \boldsymbol{\tau}_i 1_{\{i \leq t_i\}} + \boldsymbol{\eta}_t \qquad (46)$$

with $\boldsymbol{\eta}_t \sim \mathsf{IN}_k\left[\mathbf{0}, \boldsymbol{\Omega_\eta}\right]$. If the relation between $\mathbf{x}_{2,t}$ and $x_{1,t}$ is invariant, including the detected shifts should not lead to rejection in the conditional model, which can be tested by an F-test on $\boldsymbol{\tau}_i$ in (46).

The marginal model (45) is saturated by step indicators, as proposed by Castle, Doornik, Hendry, and Pretis (2015) and implemented in Autometrics (see Doornik and Hendry, 2018). Although (45) has many more candidate variables than observations, model selection based on expanding and contracting multi-path block searches has proved a viable approach following the analyses in Hendry, Johansen, and Santos (2008) and Johansen and Nielsen (2009, 2013) leading to the automatic tests of invariance in Hendry and Santos (2010) and Castle, Hendry, and Martinez (2017). The former develops impulse-indicator saturation (IIS) tests and the latter step-indicator saturation (SIS). When $k+3$ is not too large, an **explanation** of a feasible approach is the split-half analysis of IIS and SIS, where all the regressors are retained without selection and half the indicators are included and tested for significance, then replaced by the second half and finally any significant indicators are jointly included and all remaining significant ones retained. These are then included in the equation being used for the scenario study and tested for significance as in (46). Should any of the indicators that shift (45) also shift the conditional model, then invariance is rejected. In practice, Autometrics uses a multi-path block search and not just one split sample for reasons discussed by Ericsson (2017). IIS and SIS can also be applied directly to the policy equation to check its specification. The first application in Section 4 applies SIS for the level of the UK wage share, whereas the second uses IIS as it concerns growth rates, and hence the differences of steps.

4 Illustrative applications

Our two applications illustrate the top two rows of Table 1. The first relates to economics and the second to climate. In the former, the scenario variable in a mis-specified model of the UK wage share fails a test of invariance, whereas in the second, the in-sample invariant link between the scenario variable and the other conditioning variable allows the mis-specified model to better estimate the scenario difference.

4.1 Wage share example

Our first example illustrates a setting where invariance of the omitted variable with respect to shifts in the included is strongly rejected. The dependent variable is the log of the UK wage share W_t, being explained by its previous value, W_{t-1}, the unemployment rate, $U_{r,t}$, and inflation, Δp_t, where the mis-specified model (labelled as sub model) drops Δp_t. The data are from 1861 to 2006 with the observations over the Great Recession 2007–2014 retained for the scenario and forecast evaluation: Castle and Hendry (2014b) provide data sources. We consider a scenario of reducing unemployment by 2 percentage points in each year. Figure 2 plots the time series of the three variables and the scatter plot of inflation against unemployment (aka the Phillips curve–shown to be a non-constant relation in Hendry, 2015).

Table 2 records the estimates for known values of $U_{r,t}$, and Δp_t over the forecast horizon. The second column (called [1] 'Full Model'), records the estimates of the regression of the wage share W_t on a constant, W_{t-1}, $U_{r,t}$, and Δp_t, which shows a negative, but insignificant, coefficient for unemployment and a positive and highly significant one for inflation. The third column ([2] 'Sub Model') reports the estimates when Δp_t is excluded, showing a jump in the (negative) coefficient of unemployment which is now significant at 1%.[1]

As W_t and $U_{r,t}$ are levels, step-indicator saturation is the appropriate approach for detecting location shifts, and the fourth column shows the selection outcomes for the marginal model of U_r where four location shifts are detected at 0.5% significance selection to assess invariance. With 146 observations,

---

[1] *, **, *** denote significant at 5%, 1% and 0.1% respectively using conventional tests that assume well-behaved Normal errors.

Figure 2: (a) UK log wage share W_t; (b) UK rate of unemployment $U_{r,t}$ with the scenario shift shown as dashed; (c) UK rate of inflation Δp_t; (d) scatter-plot of inflation against unemployment.
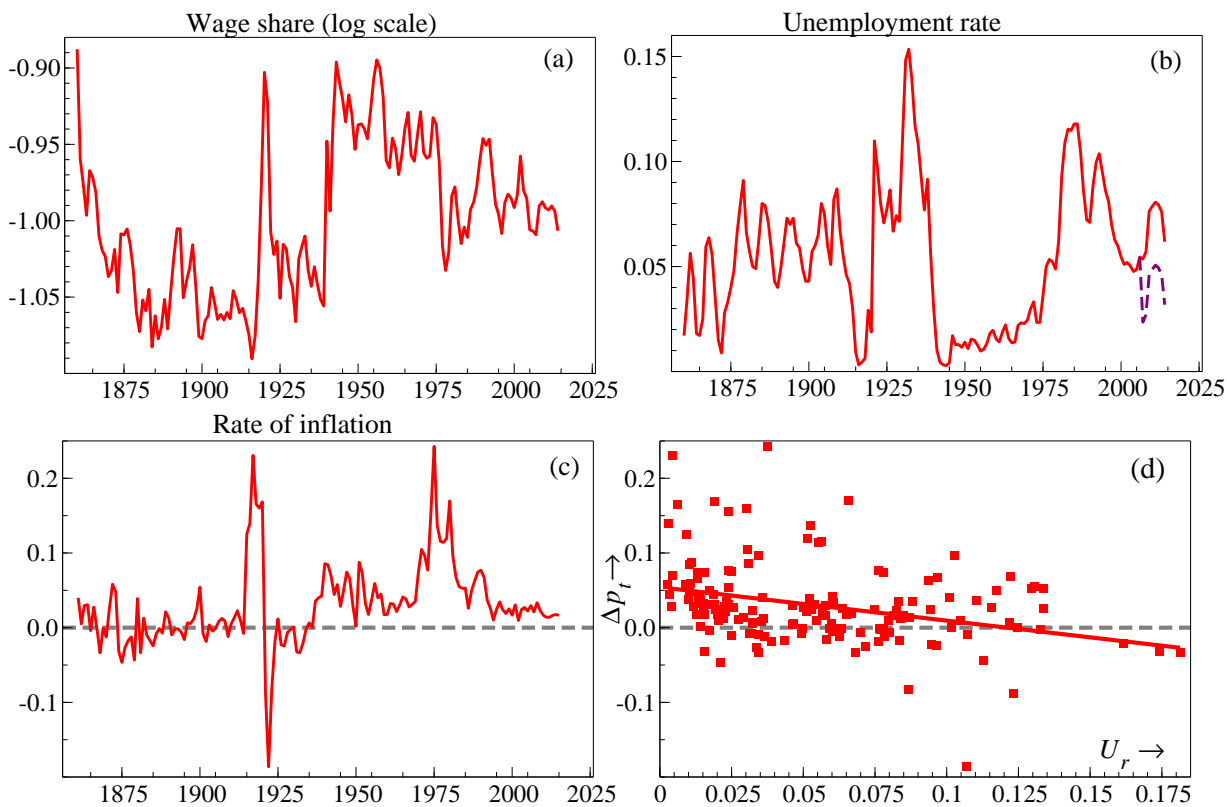
Table 2: Estimation Results: Modelling UK Wage Share W_t

Dep. Var:	W_t		$U_{r,t}$	Δp_t
	[1] Full Model	[2] Sub Model	[3] SIS	[4] testing Δp_t model
Constant	-0.17 (0.04)***	-0.16 (0.04)***	0.08 (0.003)***	0.11 (0.017)***
W_{t-1}	0.83 (0.04)***	0.83 (0.04)***	-	-
$U_{r,t}$	-0.11 (0.06)	-0.18 (0.06)**	-	-0.82 (0.19)***
Δp_t	0.11 (0.04)**	-	-	-
$S_{t=1920}$	-	-	-0.03 (0.007)***	0.04 (0.02)*
$S_{t=1929}$	-	-	-0.03 (0.010)***	-0.07 (0.02)**
$S_{t=1938}$	-	-	0.10 (0.008)***	0.02 (0.02)
$S_{t=1977}$	-	-	-0.06 (0.005)***	-0.04 (0.02)*
F-Test for S_t	-	-	[p<0.001]	[p<0.001]

16

there are 145 step indicators to select from, so on average less than one step will be significant by chance under the null ($145 \times 0.005 \approx 0.7$). Three of those four location shifts detected for unemployment are significant at 5% when entered in the marginal model for inflation on $U_{r,t}$, and the F-test of their joint relevance strongly rejects its invariance to shifts in unemployment at less than 0.1%: see Table 2 column 5. Thus, shifts to unemployment shift the relationship between unemployment and inflation (consistent with a non-constant Phillips curve), so a scenario of changing unemployment will not have the outcome anticipated from the sub-model.

Figure 3: (a) Baseline forecasts of W_t with $\pm\widehat{\sigma}_f$ bars and scenarios with $\pm\widehat{\sigma}_f$ bands from the 'full model'; (b) baseline forecasts of W_t with $\pm\widetilde{\sigma}_f$ bars and scenarios with $\pm\widetilde{\sigma}_f$ bands from the 'sub model'; (c) baseline and scenario forecasts derived from the Castle-Hendry real-wage model; (d) the three sets of scenario differences from their own baselines in (a), (b) and (c).
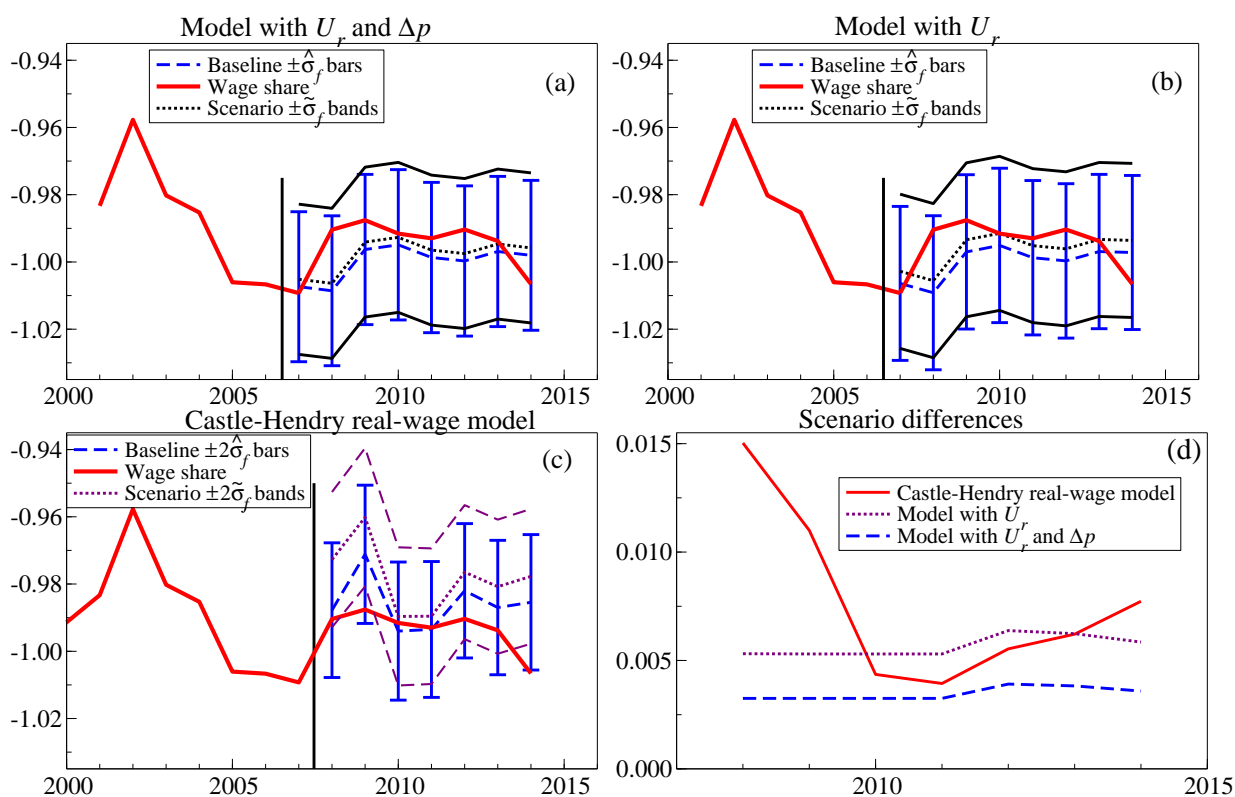
Compared to the more comprehensive model for real wages in Castle and Hendry (2014b), from which the wage share can be derived, even the 'full model' in Table 2 is badly mis-specified, with several diagnostic tests rejecting, and an error standard deviation of 2.2%, as against their model's 1.04%. As W_t is (the log of) the nominal wage bill (i.e., average wages times employment) divided by nominal output (i.e., real output times the price level), it will be affected by productivity as a key determinant as well as unemployment and inflation (both of which have non-linear effects in their model). Another test of the specification of the column 2 model is to apply SIS directly, which selects 6 step indicators at 0.1%, and now unemployment is significantly negative at 1% whereas inflation is completely insignificant.

Nevertheless, Figure 3 illustrates a scenario of lower unemployment over the Great Recession, re-

ducing it by 2 percentage points to remain close to the level of 5.4% in 2007. As the coefficients of $U_{r,t}$ in Table 2 are relatively small, the differences between scenarios and baselines for the wage share are not large in (a) and (b), with a somewhat greater increase predicted by the sub model. Both purport to correctly show a higher wage share under the scenario of lower unemployment, even though they are in fact generally lower than the actual outcome despite the large stimulus to employment. However, the difference between scenario and baseline is large for the wage share derived from the Castle–Hendry real-wage model, which encompasses most earlier models and passes tests of invariance to interventions.

4.2   Climate Impacts on Agriculture

The second example illustrates a case when invariance is not rejected empirically. Evaluation of the economic impacts of climate change commonly use scenario analyses, such as assessing economic or agricultural outcomes conditional on a particular climate path. Similarly, climate outcomes are commonly studied conditional on emission scenarios. Here we consider a simple scenario analysis of the growth in agricultural production conditional on climate variability. Burke, Hsiang, and Miguel (2015) and Pretis, Schwarz, Tang, Haustein, and Allen (2018) use the observed growth in value-added of agricultural production across countries to estimate the impacts of climate variability on agricultural outcomes across countries over time. Both temperature variability and precipitation are found to be significant factors affecting the growth in agricultural production. Here we use a single country (Indonesia) out of the panel analysis in Pretis et al. (2018) to consider a simple scenario analysis.

Figure 4: Application data for Indonesia: Growth in agricultural value-added (top-left), temperature variability (top-right), precipitation (bottom-left), and precipitation against temperature variability (bottom-right) with IIS fit.
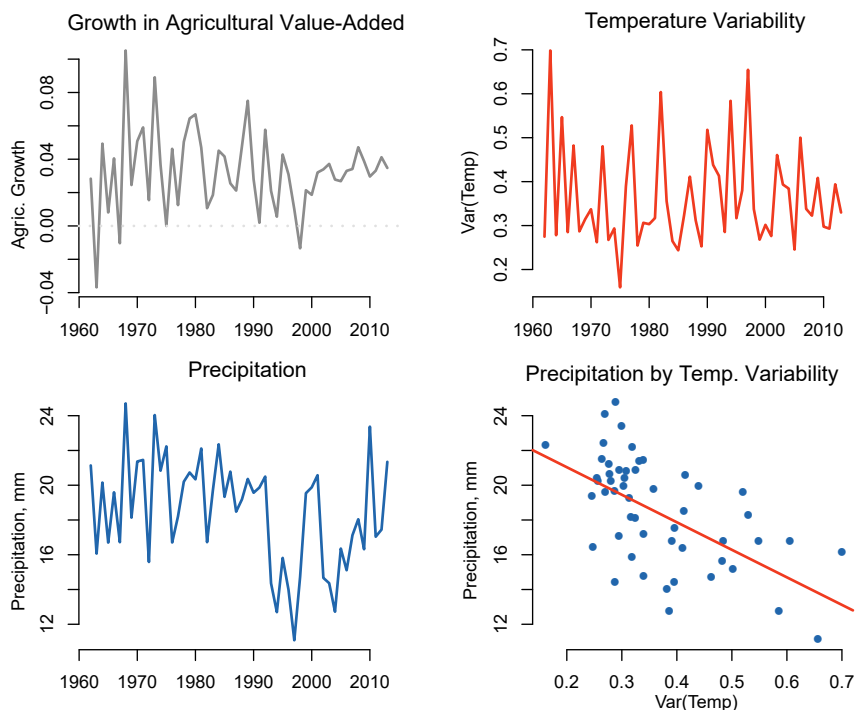
In particular, we are interested in a scenario of increasing temperature variability. This is consistent with future climate change projections of a rise in variability in the tropics (see Bathiany et al.

2018). We estimate a simple model of growth in agricultural production, and subsequently use the estimated model to study a scenario of increased temperature variability. Annual data on agricultural output growth for Indonesia is taken from the World Bank data on growth in agricultural value-added (World Bank, 2017). Climate observations of temperatures and precipitation are obtained from the Matsuura and Wilmot (2015) dataset (v.4) and mapped to countries as used in Pretis et al. (2018). Time series of the country-level data are shown in Figure 3.

We estimate the simple linear model of growth in agricultural value-added from 1962 until 2012 as a function of temperature variability and precipitation (labelled as [1] 'Full Model'), as well as a mis-specified model omitting precipitation ([2] 'Sub Model'). Model estimation results are shown in Table 3. Results of the simple model suggest a negative impact of increased temperature variability and a positive impact of precipitation.

We consider a scenario of agricultural growth impacts when temperature variability is increased by two standard deviations ($\delta_x = 0.23$). Figure 5 and Table 4 show the projected impacts and scenario differences, suggesting a reduction of agricultural growth of approximately 1.7% per annum when using the full model [1] in Table 3 which includes both temperature variability and precipitation. Using the full model [1], however, risks under-stating the projected impacts due to a potential relationship between precipitation and temperature variability. As described in Section 3, the relationship between covariates affects the accuracy of the estimated scenario difference. In the present application, there is an observed negative relationship between precipitation and temperature variability shown by [4] in Table 3 and Figure 3. While it is difficult to establish a formal causal relationship between the scenario variable and the conditioning variables, we can, however, assess whether the relationship is invariant in-sample.

To test invariance we estimate a marginal model of temperature variability (using just a constant), and detect shocks using impulse indicator saturation (IIS) – Model [3] in Table 3 (with IIS being applied at a significance level of 5% given 51 observations). The detected impulses (in the years 1963, 1982, 1994, and 1997) are then included in the conditional model of precipitation and tested for their joint significance. If the relationship between precipitation and temperature variability is invariant to interventions, then any shock to temperature variability should feed onto precipitation solely through the coefficient on temperature variability itself, and thus appear insignificant in the conditional model. An F-test for joint significance of the impulses fails to reject invariance (p = 0.41). Thus, the projected impact of an increase in temperature variability on agricultural growth likely under-states the actual impact because of the negative relationship between the scenario variable and the conditioning variable.

Table 3: Estimation Results: Modelling growth in agricultural value-added (Indonesia)

Dep. Var:	Agr.Gr.$_t$		Var(Temp)$_t$	Precip$_t$
	[1] Full Model	[2] Sub Model	[3] IIS Model	[4] Precip. Model
Constant	0.015 (0.03)	0.076 (0.01)***	0.34 (0.012)***	24.20 (1.5)***
Agr.Gr.$_{t-1}$	-0.062 (0.12)	-0.066 (0.12)	-	-
Var(Temp)$_t$	-0.072 (0.031)*	-0.11 (0.03)***	-	-15.84 (4.25)***
Precip$_t$	0.0025 (0.001)*	-	-	-
I$_{t=1963}$	-	-	0.35 (0.09)***	2.93 (2.95)
I$_{t=1982}$	-	-	0.26 (0.09)***	2.09 (2.76)
I$_{t=1994}$	-	-	0.24 (0.09)***	-2.26 (2.73)
I$_{t=1997}$	-	-	0.31 (0.09)***	-2.76 (2.86)
F-Test for I$_t$	-	-	[p <0.001]	[p = 0.41]
IIS p-value	-	-	0.05	-
Obs. T	51	51	51	51

19

Following Section 3, two approaches allow us to account for this relationship. First, the under-specified model omitting precipitation (model [2] in Table 3) implicitly captures the relationship between precipitation and agricultural growth through the bias in the estimated coefficient—the negative impact of temperature variability is larger in the model omitting precipitation. Thus, using the under-specified model to project the impact of a 2SD increase in variability leads to a larger projected negative effect of -2.6% (as seen in Figure 5) due to the omitted relationship between temperatures and precipitation.

Figure 5: Projected impact of a two standard deviation increase in temperature variability on Indonesian growth in agricultural value-added. Observations are shown in black, baseline projections shown in solid colours, scenario projections with an increase in temperature variability are shown as dashed colours. Left panel shows mis-specified model omitting precipitation. Middle panel shows projections using the full model not accounting for precipitation–temperature links (blue) as well as accounting for the link (green). Scenario differences are shown in the right panel for the under-specified model (red), the model including both temperature and precipitation but failing to account for any links (blue), and the full model estimating the link (green). Shading and error bars denote ± 1 standard error of the scenarios.
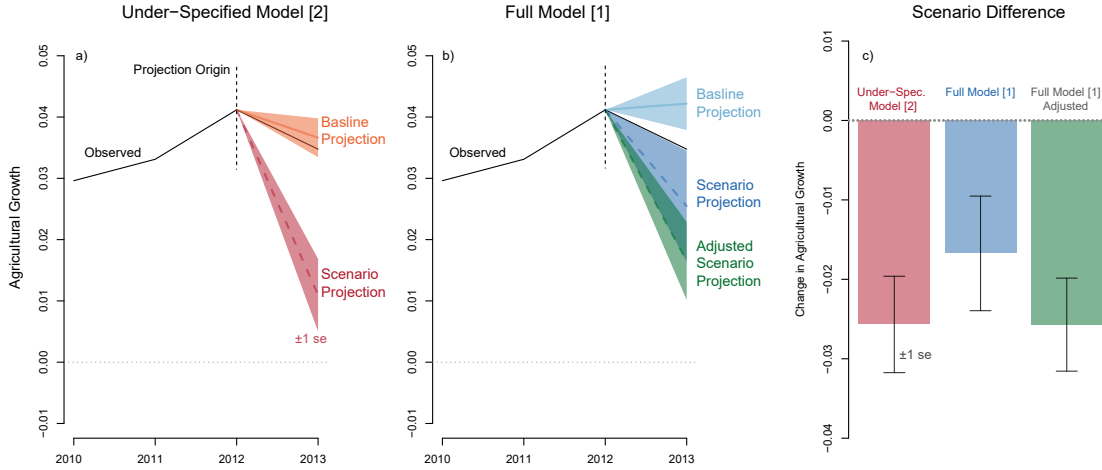
Second, the effect can be recovered by considering the joint impact of an increase in temperature variability using the estimated model linking precipitation and temperatures [4]. A two standard deviation increase in temperature variability has the estimated effect of -1.7% from [1] plus the indirect effect through precipitation which can be recovered using equation [4] and subsequently included in the projection using [1]. The combined projection in Table 4 taking the relationship between conditioning variable and policy variable into account matches that of the under-specified model with an impact of -2.6% (estimated standard errors in parentheses).

Table 4: Scenario Projections for a $2\widehat{\sigma}$ Increase in Var(Temp)

Projection	Full Model [1]	Sub Model [2]	Full Model [1] (Adjusted)
$\widehat{y}_{T+1\|T}$	0.042 (0.004)	0.036 (0.003)	0.042 (0.004)
$\widehat{y}^*_{T+1\|T}$	0.025 (0.009)	0.011 (0.006)	0.016 (0.006)
$\widehat{y}^*_{T+1\|T} - \widehat{y}_{T+1\|T}$	-0.017 (0.007)	-0.026 (0.006)	-0.026 (0.006)

Consequently, an under-specified model can provide accurate estimates of a policy impact in some settings, but not in others. In both cases, if the data are available, careful modelling of the links between

covariates is required to establish whether the relationships are indeed invariant to interventions or merely in-sample correlations that will prove to be unstable following interventions. Such results for scenarios are closely similar to the findings in Castle and Hendry (2014a) for selecting models in DGPs with breaks.

5   Conclusion

We considered inference on the difference between scenarios when using estimated models of an unknown data generation process (DGP). For univariate models, our results show that when the model is linear in the single variable perturbed by the scenario intervention of a deterministic shift, then the significance of the difference between a scenario intervention and a baseline of no intervention depends only on the significance of the in-sample t-test of that variable. Under these circumstances, a scenario comparison adds little beyond testing for the significance of the variable in the estimated model. When models are more general and include multiple covariates together with a variable perturbed by a scenario analysis, then inference on the scenario difference crucially depends on the relationship between the conditioning variables and the policy variable. If there exists an invariant causal relationship between variables, a simple model omitting control variables may be preferred over a well-specified model that matches the DGP, but not otherwise. These results were illustrated by two empirical examples, the reaction of the wage share to a reduction in unemployment and of an increase in temperature variability on agricultural growth, but apply more widely to scenarios of the health and economic impacts from the SARS-CoV-2 virus that causes COVID-19.

The recommendations are thus mixed—an estimated conditional model matching its DGP equation does not automatically yield an unbiased estimate of the true outcome. Only if there is no relationship between the covariates and the perturbed variable is the scenario analysis informative, unless the relationships between conditioning variables and policy variable are explicitly modelled. If there exists an invariant causal relationship between the conditioning variables and the policy variable, then a mis-specified model omitting additional covariates may be preferred, similar to the problem of conditioning on a post-treatment variable in causal inference, or deliberately omitting exogenous variables from an open forecasting model. However, if the relationship between covariates and the policy variable is merely an in-sample projection that is not invariant to an intervention, then the resulting parameter changes following a scenario intervention can render that scenario analysis uninformative.

Thus, although a model-based scenario is unique, the 'real-world outcome' is not, and will vary in unknown ways with unknown omitted influences and parameter changes. The recommended approach in Hendry and Doornik (2014) of seeking a large information set to nest the DGP, selecting the significant effects therefrom and testing for parameter invariance to past large changes would help avoid problems of mis-specification and non-invariance by selecting a model that was a good approximation to the local DGP. Nevertheless, there are situations noted above when the model is mis-specified for the DGP but the scenario difference is correct: omitting variables from linear models will bias parameter estimates but may not distort calculations of scenario differences and forecasts.

References

Bathiany, S., V. Dakos, M. Scheffer, and T. M. Lenton (2018). Climate models predict increasing temperature variability in poor countries. *Science Advances 4*, eaar5809.

Burke, M., S. M. Hsiang, and E. Miguel (2015). Global non-linear effect of temperature on economic production. *Nature 527*, 235–239.

Castle, J. L., J. A. Doornik, D. F. Hendry, and F. Pretis (2015). Detecting location shifts during model selection by step-indicator saturation. *Econometrics 3(2)*, 240–264.

Castle, J. L. and D. F. Hendry (2014a). Model selection in under-specified equations facing breaks. *Journal of Econometrics 178(2)*, 286–293.

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

Castle, J. L., D. F. Hendry, and A. B. Martinez (2017). Evaluating forecasts, narratives and policy using a test of invariance. *Econometrics 5(39)*, DOI:10.3390/econometrics5030039.

Chevillon, G. and D. F. Hendry (2005). Non-parametric direct multi-step estimation for forecasting economic processes. *International Journal of Forecasting 21*, 201–218.

Clements, M. P. and D. F. Hendry (1998). *Forecasting Economic Time Series*. Cambridge: Cambridge University Press.

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

Engle, R. F., D. F. Hendry, and J.-F. Richard (1983). Exogeneity. *Econometrica 51*, 277–304.

Ericsson, N. R. (2017). Interpreting estimates of forecast bias. *International Journal of Forecasting 33*, 563–568.

Flyvbjerg, B., M. K. S. Holm, and S. L. Buhl (2005). How (in)accurate are demand forecasts in public works projects?: The case of transportation. *Journal of the American Planning Association 71*(2), 131–146.

Hendry, D. F. (2015). *Introductory Macro-econometrics: A New Approach*. London: Timberlake Consultants.

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

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

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

Hendry, D. F. and G. E. Mizon (2012). Open-model forecast-error taxonomies. In X. Chen and N. R. Swanson (Eds.), *Recent Advances and Future Directions in Causality, Prediction, and Specification Analysis*, pp. 219–240. New York: Springer.

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

Johansen, S. and B. Nielsen (2009). An analysis of the indicator saturation estimator as a robust regression estimator. In J. L. Castle and N. Shephard (Eds.), *The Methodology and Practice of Econometrics*, pp. 1–36. Oxford: Oxford University Press.

Johansen, S. and B. Nielsen (2013). Outlier detection in regression using an iterated one-step approximation to the Huber-skip estimator. *Econometrics 1*, 53–70.

Matsuura, K. and C. J. Willmott (2015). Terrestrial air temperature and precipitation: Monthly and annual time series (1900–2014) v. 4.01. http://climate.geog.udel.edu/ climate/html_pages/global2014/readme.globaltsp2014.html.

Pretis, F., M. Schwarz, K. Tang, K. Haustein, and M. R. Allen (2018). Uncertain impacts on economic growth when stabilizing global temperatures at 1.5°C or 2°C warming. *Philosophical Transactions of the Royal Society A376: 20160460*. https://doi.org/10.1098/rsta.2016.0460.

Weale, M. (2016). Brexit and monetary policy, Bank of England. https://www.bankofengland.co.uk/speech/2016/brexit-and-monetary-policy.

Webster, M. D., M. Babiker, M. Mayer, J. M. Reilly, J. Harnisch, M. C. Sarofim, and C. Wang (2002). Uncertainty in emissions projections for climate models. *Atmospheric Environment 36*(22), 3659–3670.

Webster, M. D., C. Forest, J. Reilly, M. Babiker, D. Kicklighter, M. Mayer, R. Prinn, M. Sarofim, A. Sokolov, P. Stone, and C. Wang (2003). Uncertainty analysis of climate change and policy response. *Climatic change 61*(3), 295–320.

World Bank (2017). World development indicators, WDI. http://data.worldbank.org/products/wdi.

Appendix: Approximate variances of multi-period scenario outcomes

To illustrate for the model with a single policy variable in (3), let $\boldsymbol{\theta} = (\beta_1 : \lambda)'$ so the scenario path depends on $f(\widehat{\boldsymbol{\theta}})$ for which we need $\mathsf{V}[f(\widehat{\boldsymbol{\theta}})]$. Using a linear expansion where $\boldsymbol{\xi} = (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})$:

$$f\left(\widehat{\boldsymbol{\theta}}\right) = f(\boldsymbol{\theta} + \boldsymbol{\xi}) \approx f(\boldsymbol{\theta}) + \frac{\partial f(\boldsymbol{\theta})'}{\partial \boldsymbol{\theta}} \boldsymbol{\xi} = f(\boldsymbol{\theta}) + \frac{\partial f(\boldsymbol{\theta})'}{\partial \boldsymbol{\theta}} \left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right) \quad (47)$$

with from (6):

$$\widehat{\boldsymbol{\theta}} \sim \mathsf{N}_2\left[\boldsymbol{\theta}, \mathsf{V}\left[\widehat{\boldsymbol{\theta}}\right]\right] = \mathsf{N}_2\left[\begin{pmatrix} \beta_1 \\ \lambda \end{pmatrix}, T^{-1}\sigma_\epsilon^2 \begin{pmatrix} \sigma_{11}^{-1} & 0 \\ 0 & \sigma_{22}^{-1} \end{pmatrix}\right]$$

so:

$$\mathsf{V}\left[f\left(\widehat{\boldsymbol{\theta}}\right)\right] \approx \frac{\partial f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \mathsf{V}\left[\widehat{\boldsymbol{\theta}}\right] \frac{\partial f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}$$

which can be generalized depending on the model and scenario under analysis. When:

$$\mathsf{E}\left[\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right)\right] = \mathbf{0}$$

then (47) also entails that:

$$\mathsf{E}\left[f\left(\widehat{\boldsymbol{\theta}}\right)\right] \approx f\left(\mathsf{E}\left[\widehat{\boldsymbol{\theta}}\right]\right) \quad (48)$$

so using (48):

$$\mathsf{E}\left[\left(\widehat{y}_{T+n|T}^* - \widehat{y}_{T+n|T}\right) \mid x_{1,T+n}, \ldots, x_{1,T+1}, y_T\right] = \mathsf{E}\left[\widehat{\beta}_1\left(\sum_{i=0}^{n-1}\widehat{\lambda}^i\right)\delta_x\right] \approx \beta_1\left(\sum_{i=0}^{n-1}\lambda^i\right)\delta_x \quad (49)$$

As the scenario difference here depends on $f\left(\widehat{\boldsymbol{\theta}}\right) = \widehat{\beta}_1(\sum_{i=0}^{n-1}\widehat{\lambda}^i)$, we have:

$$\frac{\partial f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} = \frac{\partial \beta_1\left(\sum_{i=0}^{n-1}\lambda^i\right)}{\partial (\beta_1 : \lambda)'} = \left(\left(\sum_{i=0}^{n-1}\lambda^i\right) : \beta_1\left(\sum_{i=1}^{n-1}i\lambda^{i-1}\right)\right) \quad (50)$$

so:

$$\mathsf{V}\left[f\left(\widehat{\boldsymbol{\theta}}\right)\right] \approx T^{-1}\sigma_\epsilon^2 \left(\left(\sum_{i=0}^{n-1}\lambda^i\right) : \beta_1\left(\sum_{i=1}^{n-1}i\lambda^{i-1}\right)\right)\left(\begin{array}{cc}\sigma_{11}^{-1} & 0 \\ 0 & \sigma_{22}^{-1}\end{array}\right)\left(\begin{array}{c}\left(\sum_{i=0}^{n-1}\lambda^i\right) \\ \beta_1\left(\sum_{i=1}^{n-1}i\lambda^{i-1}\right)\end{array}\right)$$

$$= T^{-1}\sigma_\epsilon^2\left(\left(\sum_{i=0}^{n-1}\lambda^i\right)^2\sigma_{11}^{-1} + \beta_1^2\sigma_{22}^{-1}\left(\sum_{i=1}^{n-1}i\lambda^{i-1}\right)^2\right) \tag{51}$$

Thus, the variance, $\mathsf{V}[\widehat{y}^*_{T+n|T} - \widehat{y}_{T+n|T}]$ of the estimator $\left(\widehat{y}^*_{T+n|T} - \widehat{y}_{T+n|T}\right)$ is then approximately:

$$T^{-1}\sigma_\epsilon^2\left(\left(\sum_{i=0}^{n-1}\lambda^i\right)^2\sigma_{11}^{-1} + \beta_1^2\left(\sum_{i=1}^{n-1}i\lambda^{i-1}\right)^2\sigma_{22}^{-1}\right)\delta_x^2$$

and as expected, will be larger than at 1-step, reflecting the dynamics. Using:

$$\sum_{i=0}^{n-1}\lambda^i = \frac{1-\lambda^n}{1-\lambda} \text{ and } \sum_{i=1}^{n-1}i\lambda^{i-1} = \frac{1-\left(n\lambda^{n-1}\left(1-\lambda\right)+\lambda^n\right)}{\left(1-\lambda\right)^2}$$

then $\mathsf{V}\left[\widehat{y}^*_{T+n|T} - \widehat{y}_{T+n|T}\right]$ is approximately:

$$\frac{\sigma_\epsilon^2\left(\left(1-\lambda^n\right)^2\left(1-\lambda\right)^2\sigma_{11}^{-1} + \beta_1^2\left(\left(1-\left(n\lambda^{n-1}\left(1-\lambda\right)+\lambda^n\right)\right)^2\right)\sigma_{22}^{-1}\right)\delta_x^2}{T\left(1-\lambda\right)^4} \tag{52}$$