Identification in (Endogenously) Nonlinear SVARs is Easier than you Think

James A. Duffy*
University of Oxford

Sophocles Mavroeidis†
University of Oxford

March 2026

Abstract

We study identification in structural vector autoregressions (SVARs) in which the endogenous variables enter nonlinearly on the left-hand side of the model, a feature we term endogenous nonlinearity, to distinguish it from the more familiar case in which nonlinearity arises only through exogenous or predetermined variables. This class of models accommodates asymmetric impact multipliers, endogenous regime switching, and occasionally binding constraints. Our main result shows that, under weak regularity conditions, the model parameters are identified from the data up to, and only up to, an orthogonal transformation, precisely as in the linear SVARs. This nonparametric identification result has the powerful implication that most existing identification schemes for linear SVARs extend directly to the nonlinear setting, with the number of restrictions required for exact identification unchanged. We specialise these results to piecewise affine SVARs and propose a kernel-convolution approach to smooth transitions that preserves invertibility of the structural form. We illustrate the methodology with an application to a nonlinear Phillips curve, showing that the presence of nonlinearity can be tested robustly across all identifying assumptions, and finding significant evidence of state-dependent inflation dynamics.

Theorem 2.2 supersedes a result that first appeared, with rather stronger assumptions, as Theorem A.1 in Duffy and Mavroeidis (2024, v2).

*Department of Economics and Corpus Christi College; james.duffy@economics.ox.ac.uk
†Department of Economics and University College; sophocles.mavroeidis@economics.ox.ac.uk

Contents

1 Introduction **1**

2 Observational equivalence and identification **4**
2.1 The linear SVAR: a brief review 4
2.2 The (endogenously) nonlinear SVAR 6
2.3 Orthogonal reduced-form parametrisation 10

3 Piecewise affine SVARs **11**
3.1 Endogenous regime switching 11
3.2 Identification 12
3.3 Smooth transitions 15

4 Application: a nonlinear Phillips curve? **17**
4.1 Formulation as an endogenous regime-switching SVAR 17
4.2 Testing for linearity in the Phillips curve 18
4.3 Results 19

A Proofs of identification results **24**
A.1 Reformulation of the problem 24
A.2 Preliminaries 25
A.3 Proof of Theorem A.1 26
A.4 Proof of Theorem 2.2 32

B Proofs for piecewise affine functions **33**
B.1 Proof of Proposition 3.1 33
B.2 Proof of Lemma B.1 36

1 Introduction

For more than four decades, following the seminal contribution of Sims (1980), the linear structural vector autoregression (SVAR)

$$\Phi_0 z_t = c + \sum_{i=1}^{k} \Phi_i z_{t-i} + \varepsilon_t, \qquad \varepsilon_t \sim_{\text{i.i.d.}} [0, I_p], \tag{1.1}$$

has played a central role in empirical macroeconomics. This is a dynamic linear simultaneous equations model (SEM), in which the p endogenous variables z_t are jointly determined as a function of their past values and the p (mutually uncorrelated) structural shocks ε_t. The latter are regarded as the exogenous inputs to the system, so that causality is understood to run from these shocks to current and future values of z_t, and a key object of interest is the mapping between ε_t and z_{t+h} for $h \geq 0$: the (structural) impulse response function.

In this context, a fundamental result that characterises the extent to which the data is informative about the model parameters, and thus also about those impulse responses, may be phrased heuristically as follows:

> (ID) Data on $\{z_t\}$ is sufficient to identify the linear SVAR parameters $(c, \{\Phi_i\}_{i=0}^p)$, and the structural shocks ε_t, up to, and only up to, an orthogonal matrix Q.

In light of this, what might be termed the 'SVAR identification problem' becomes one of finding sufficient additional restrictions on that matrix Q, so as to pin down, wholly or partially, the model parameters. The literature since has developed a variety of ways of using macroeconomic theory to generate such restrictions, based e.g. on the relative timing of shocks, the signs of their effects on impact, their medium- and long-run effects, and their correlation with external instruments (for a textbook discussion of which, see Kilian and Lütkepohl, 2017).

The linearity of (1.1) is convenient, but inherently limiting as to the nature of the dynamics that can be modelled. In particular, it has the rather undesirable implication that the response of the economy to shocks must be the same *irrespective* of the phase of the business cycle: so that e.g. an aggregate demand shock has exactly the same effect on unemployment and inflation in the depths of a recession, when there is considerable slack in the labour market, as it would during periods of expansion. The substantial literature on nonlinear (S)VAR models has arisen partly to address these limitations (see e.g. Chan, 2009; Teräsvirta, Tjøstheim, and Granger, 2010; Hubrich and Teräsvirta, 2013, for surveys). These allow the parameters of the SVAR at time t to depend on an exogenous (or if not wholly exogenous, at least *predetermined*) switching process s_{t-1}, as e.g. in[1]

$$\Phi_0(s_{t-1}) z_t = c(s_{t-1}) + \sum_{i=1}^{k} \Phi_i(s_{t-1}) z_{t-i} + \varepsilon_t, \tag{1.2}$$

where often $s_t \in \{1, \ldots, L\}$ takes finitely many values, and each $\Phi_i(s) = \sum_{\ell=1}^{L} \pi_\ell(s) \Phi_i^{(\ell)}$ switches, or smoothly transitions, between the parameters of the L 'regimes'; here each $\pi_\ell(s) \in [0, 1]$, with

[1] In many treatments of this model, the regime indicator in (1.2) is denoted as s_t, rather than s_{t-1}. However, a feature of these models is that the regime is always determined prior to the realisation of ε_t, and may thus be regarded as measurable with respect to time-$(t-1)$ information; we have written s_{t-1} to make this clearer.

$\sum_{\ell=1}^{L} \pi_\ell(s) = 1$. The evolution of $\{s_t\}$ may be modelled as an exogenous Markov chain (as in a Markov switching model), possibly with state-dependent transition probabilities, or as a function of some predetermined variables (such as an element of z_{t-i} for some $i \geq 1$, as in a typical 'threshold autoregressive' model); but in any case, s_{t-1} must be determined prior to the realisation of ε_t. We therefore refer to these henceforth as *exogenous* regime-switching SVARs. (This characterisation applies to time-varying parameter VARs, in which $\{s_t\}$ is also some exogenous but possibly nonstationary process, such as a random walk.)

While models of the form (1.2) enjoy greatly enriched dynamics relative to (1.1), here the possibility of regime switching exacerbates the identification problem. Indeed the counterpart of (ID) for general Markov-switching models is that, *conditional on the regime* $s_{t-1} = s$, the parameters of (1.2) are identified up to an orthogonal matrix $Q(s)$. Since this matrix may vary with $s \in \{1, \ldots, L\}$, the number of unidentified parameters, and thus the number of restrictions needed to deliver (exact) identification, scales proportionally with the number of states L. In practice, this may necessitate replicating a common set of $p(p-1)/2$ restrictions across *all* L regimes (see e.g. Rubio-Ramirez, Waggoner, and Zha, 2005, Sec. II; Sims and Zha, 2006, Sec. III), yielding $Lp(p-1)/2$ restrictions in total. Similarly, in their two-regime STVAR model, Auerbach and Gorodnichenko (2012, p. 4) impose a Cholesky ordering on the elements of z_t in each regime.

The exogeneity of the regime (i.e. s_{t-1}) moreover restricts the kinds of nonlinearities that may be exhibited by the model's impulse responses. Notably, since each regime is itself a linear SVAR, the immediate effects of the shocks (i.e. the impact multipliers) must be *linear* in ε_t: which in particular rules out the possibility of sign-dependent asymmetries. It also renders (1.2) unable to accommodate occasionally binding constraints, such as the zero lower bound (ZLB) constraint on short-term nominal interest rates, because the model requires the regime (whether 'constrained' or 'unconstrained') to be determined *prior* to that of the potentially constrained variable – whereas, as a matter of logic, it ought to be the value of that variable which determines whether the model is in fact in the constrained or unconstrained regime (see Aruoba, Mlikota, Schorfheide, and Villalvazo, 2022).

Recently, Mavroeidis (2021) and Aruoba et al. (2022) introduced the first SVAR models involving what we here refer to as *endogenous* regime switching, which are notably distinguished from the earlier literature on the nonlinear SVARs of the form (1.2) in that they permit the autoregressive 'regime' to be determined *jointly* with the values of the endogenous variables. For example, the 'censored and kinked SVAR' (CKSVAR) of Mavroeidis (2021) takes the form

$$\phi_0^+ y_t^+ + \phi_0^- y_t^- + \Phi_0^x x_t = c + \sum_{i=1}^{k} [\phi_i^+ y_{t-i}^+ + \phi_i^- y_{t-i}^- + \Phi_i^x x_{t-i}] + \varepsilon_t.$$

where y_t^+ and y_t^- denote the positive and negative parts of y_t (a scalar), and x_t is $(p-1)$-dimensional. In this model there are two contemporaneous regimes: one associated with $y_t > 0$ (the 'unconstrained' regime, in the ZLB setting), and the other with $y_t \leq 0$ (the 'constrained' regime), and in every period the model is solved *simultaneously* for the current values of y_t and x_t, and for the applicable regime (as depends on the sign of y_t). Thus in situations where the $\varepsilon_t = 0$ would entail a solution of $y_t = 0$ (or approximately so), this allows the impact multipliers

of ε_t to be asymmetric, depending on which regime they push the model in to.

Building on these developments, this paper proposes a new class of nonlinear SVAR models, which have the general form

$$f_0(z_t) = \sum_{i=1}^{k} f_i(z_{t-i}) + \varepsilon_t \tag{1.3}$$

where each $f_i : \mathbb{R}^p \to \mathbb{R}^p$ is a continuous, possibly nonlinear function, with f_0 being invertible; we refer to these models as being 'endogenously nonlinear', in view of the nonlinearities on the l.h.s. Because it is not tied to any particular functional form, (1.3) also offers a great deal of flexibility in its dynamics, comparable to that offered by (1.2). This framework readily encompasses the CKSVAR, which corresponds to a special case in which each f_i is piecewise linear. More general models with endogenous switching between several regimes may be straightforwardly encompassed within the framework (1.3), by specifying

$$f_0(z) = \sum_{\ell=1}^{L} \mathbf{1}\{z \in \mathscr{Z}^{(\ell)}\}(\bar{\phi}_0^{(\ell)} + \Phi_0^{(\ell)} z) \tag{1.4}$$

to be an invertible, (continuous) piecewise affine function, where $\{\mathscr{Z}^{(\ell)}\}_{\ell=1}^{L}$ is a convex partition of $\mathbb{R}^p$, and the current regime ℓ_t corresponds to the element of that partition for which $z_t \in \mathscr{Z}^{(\ell_t)}$.

The principal contribution of this paper is to characterise observational equivalence in the setting of the following, slightly more general formulation of (1.3),

$$f_0(z_t) = \boldsymbol{f}_1(\boldsymbol{z}_{t-1}) + \varepsilon_t, \qquad \varepsilon_t \sim_{\text{i.i.d.}} [0, I_p], \tag{1.5}$$

where $\boldsymbol{z}_{t-1} = (z_{t-1}^{\mathsf{T}}, \ldots, z_{t-k}^{\mathsf{T}})^{\mathsf{T}}$, and $\boldsymbol{f}_1 : \mathbb{R}^{kp} \to \mathbb{R}^p$ need not be separable in the lags of z_t (Section 2). Remarkably, despite the far greater flexibility afforded by the parametrisation of (1.5) relative to the linear SVAR (1.1), the fundamental identification result (ID) carries over to (1.5) essentially unchanged. Under weak conditions on the functions $(f_0, \boldsymbol{f}_1)$ and the distribution of the shocks, we have (Theorem 2.2):

(ID′) Data on $\{z_t\}$ is sufficient to identify the nonlinear SVAR parameters $(f_0, \boldsymbol{f}_1)$, and the structural shocks ε_t, up to, and only up to, an orthogonal matrix Q.

This is a *nonparametric* identification result, in the sense that we do not suppose that $(f_0, \boldsymbol{f}_1)$ have any particular (known) parametric form. While its proof draws upon the microeconometrics literature on nonlinear SEMs (see in particular Matzkin, 2008, 2015; Berry and Haile, 2018; Chernozhukov, Galichon, Henry, and Pass, 2021), it constitutes a genuinely novel result within that setting. (ID′) has the powerful implication that most of the existing identification results for linear SVARs apply directly to the endgenously nonlinear SVAR, since in both cases exact identification is a matter of imposing $p(p-1)/2$ restrictions sufficient to pin down Q.

There follows a discussion of the L-regime endogenous regime-switching SVAR, which arises when f_0 is specified to have the piecewise affine form (1.4), and of how to verify the conditions of Theorem 2.2 in this case (Section 3). (Here we also suppose, mostly to provide a practically convenient parametrisation, that the SVAR has the time-separable form (1.3), with each $\{f_i\}_{i=1}^{k}$

also specified to have the same functional form as (1.4).) In this context, our results imply that it is sufficient, for the purposes of exact identification, to impose identifying restrictions in *only one of those L regimes*, or even to distribute these in some way across those regimes. To obtain smooth transitions between adjacent regimes, we propose to convolve f_0 with a smooth kernel. This has the considerable advantage of preserving the invertibility of f_0, whereas this may fail if one attempts to smooth f_0 by the usual device of replacing each indicator function in (1.4) by a smooth, cdf-like function (as is very commonly done to produce 'smooth transition' (S)VARs).

Our methodology is illustrated by estimating an endogenously regime-switching SVAR (in the log vacancy–unemployment ratio and inflation), to investigate the possibility of a nonlinear Phillips curve relationship (Section 4) that was recently proposed by Benigno and Eggertsson (2023) to explain the recent post-pandemic inflation surge. In particular, the identification result given above can help investigate the robustness of the evidence of nonlinearity to alternative identification assumptions, thus shedding light on the recent debate about the importance of such nonlinearity between Benigno and Eggertsson (2023) and Beaudry, Hou, and Portier (2025).

Notation. $e_{m,i}$ denotes the ith column of an $m \times m$ identity matrix; when m is clear from the context, we write this simply as e_i. $\|\cdot\|$ denotes the Euclidean norm on $\mathbb{R}^m$. Matrix norms are always those induced by the corresponding vector norm. For a function $g : \mathbb{R}^m \to \mathbb{R}^n$, $Dg(u_0) = [(\partial g_i/\partial u_j)(u_0)]$ denotes the $(n \times m)$ Jacobian (matrix) of $g(u)$ at $u = u_0$.

2 Observational equivalence and identification

2.1 The linear SVAR: a brief review

Our point of departure is the linear SVAR, in which the observed series $\{z_t\}$ are regarded as being generated linearly from an underlying p-dimensional sequence of *structural shocks* $\{\varepsilon_t\}$, each of which have an economic interpretation (as e.g. an aggregate supply shock, a monetary policy shock, etc.) That is, for some $k \in \mathbb{N}$,

$$\Phi_0 z_t = c + \sum_{i=1}^{k} \Phi_i z_{t-i} + \varepsilon_t =: c + \mathbf{\Phi}_1 \boldsymbol{z}_{t-1} + \varepsilon_t \tag{2.1}$$

where z_t and ε_t are $\mathbb{R}^p$-valued, and to permit a more compact presentation we have defined $\mathbf{\Phi}_1 \coloneqq [\Phi_1, \ldots, \Phi_k]$ and $\boldsymbol{z}_{t-1} \coloneqq (z_{t-1}^\mathsf{T}, \ldots, z_{t-k}^\mathsf{T})^\mathsf{T}$.

Observational equivalence in this setting being well understood (see e.g. Hamilton, 1994, Ch. 11; Lütkepohl, 2007, Ch. 9), our purpose here is to briefly review this in a manner that facilitates the comparison with our results for the endogenously nonlinear SVAR, to be introduced below. To simplify the problem, we suppose that $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is i.i.d., with a (Lebesgue) density $\varrho \in \mathscr{R}$, normalised to have $\mathbb{E}\varepsilon_t = 0$ and $\mathbb{E}\varepsilon_t\varepsilon_t^\mathsf{T} = I_p$. By the Markov property the joint density of $\{z_t\}_{t=1}^T$, conditional on $\boldsymbol{z}_0$, is simply the product of the conditional densities of $z_t \mid \boldsymbol{z}_{t-1}$, for $t \in \{1, \ldots, T\}$. Under our assumptions, this (Lebesgue) density is time-invariant, and equal to

$$\varphi_{z_t|\boldsymbol{z}_{t-1}}(\xi \mid \boldsymbol{\xi}_{-1}) = \varrho(\Phi_0\xi - \mathbf{\Phi}_1\boldsymbol{\xi}_{-1}) \cdot |\det \Phi_0|,$$

where $\xi \in \mathbb{R}^p$, $\boldsymbol{\xi}_{-1} \in \mathbb{R}^{kp}$. Accordingly, we say that two alternative parameterisations of the

linear SVAR, $(c, \Phi_0, \mathbf{\Phi}_1, \varrho)$ and $(\tilde{c}, \tilde{\Phi}_0, \tilde{\mathbf{\Phi}}_1, \tilde{\varrho})$, are *observationally equivalent* if they imply identical conditional densities $f_{z_t|\boldsymbol{z}_{t-1}}$; in which case they also yield identical (conditional) likelihoods, for every possible realisation of $\{z_t\}$.

We then have the following well-known result, that observation of $\{z_t\}$ identifies the SVAR coefficients $(c, \Phi_0, \mathbf{\Phi}_1)$ up to, and only up to, an orthogonal transformation. Let $\mathbb{O}(p)$ denote the set of $p \times p$ orthogonal matrices.

Theorem 2.1. *Let $(\tilde{c}, \tilde{\Phi}_0, \tilde{\mathbf{\Phi}}_1) \in \mathbb{R}^p \times \mathbb{R}^{p\times p} \times \mathbb{R}^{p\times kp}$. Then there exists a $\tilde{\varrho} \in \mathscr{R}$ such that $(\tilde{c}, \tilde{\Phi}_0, \tilde{\mathbf{\Phi}}_1, \tilde{\varrho})$ is observationally equivalent to $(c, \Phi_0, \mathbf{\Phi}_1, \varrho)$ in the model (2.1), if and only if there exists a $Q \in \mathbb{O}(p)$ such that*

$$\tilde{c} = Qc \qquad \tilde{\Phi}_0 = Q\Phi_0 \qquad \tilde{\mathbf{\Phi}}_1 = Q\mathbf{\Phi}_1.$$

Remark 2.1. (i). Versions of this result, or equivalent characterisations thereof, have long been utilised in the analysis of linear SVARs, and linear simultaneous equations models (SEMs). This particular characterisation leads naturally to the 'orthogonal reduced-form parametrisation' (Arias, Rubio-Ramirez, and Waggoner, 2018, Sec. 2.3) of the SVAR, in terms of the (unidentified) $Q \in \mathbb{O}(p)$ and the (identified) reduced form parameters ($\Phi_0^{-1}\mathbf{\Phi}_1$ and $\Phi_0^{\intercal}\Phi_0$), which has proved fruitful for the analysis of sign-restricted SVARs (Faust, 1998; Uhlig, 1998, 2005; Arias et al., 2018), and the formulation of rank conditions for global identification (Rubio-Ramirez, Waggoner, and Zha, 2010).

(ii). The preceding follows as a corollary to Theorem 2.2 below, albeit that result is proved under stronger regularity conditions on the allowable set of densities $\mathscr{R}$. Because of the linearity of (2.1), the same result also holds under weaker conditions on the model than we have maintained here. For example, we could require $\{\varepsilon_t\}$ merely to be stationary white noise, since all that is really needed to identify the reduced form parameters is the orthogonality of ε_t from $\boldsymbol{z}_{t-1}$. On the other hand, the assumption that $\{\varepsilon_t\}$ is an i.i.d. process, often with a *known* (often Gaussian) distribution is common in empirical work, particularly in the context of Bayesian SVARs, and even in discussions of identification in these models (as in e.g. Rubio-Ramirez et al., 2010).

(iii). Here we have maintained only that the individual elements of $\varepsilon_t = (\varepsilon_{1t}, \ldots, \varepsilon_{pt})^{\intercal}$ are contemporaneously *orthogonal*, rather than being independent. We thereby exclude the possibility, highlighted in a strand of the linear SVAR literature (e.g. Lanne, Meitz, and Saikkonen, 2017; Gouriéroux, Monfort, and Renne, 2020), of exploiting the additional restrictions available when the shocks are independent *and* non-Gaussian, to strengthen the above result to one in which the SVAR coefficients are identified up to an unknown (signed) permutation matrix.

(iv). Let k_0 denote the true lag order of the SVAR, i.e. the greatest $i \in \mathbb{N}$ such that $\Phi_i \neq 0$. We have implicitly maintained that this is less than or equal to k, which may therefore be interpreted as a upper bound on the true lag order of the model. In this sense, Theorem 2.1 does not assume knowledge of the true lag order k_0 of the SVAR, but merely of some finite upper bound k thereof.

2.2 The (endogenously) nonlinear SVAR

We now seek to extend Theorem 2.1 to the setting of the following (endogenously) *nonlinear* SVAR

$$f_0(z_t) = \boldsymbol{f}_1(\boldsymbol{z}_{t-1}) + \varepsilon_t \tag{2.2}$$

where $f_0 : \mathbb{R}^p \to \mathbb{R}^p$ is invertible, and $\boldsymbol{f}_1 : \mathbb{R}^{kp} \to \mathbb{R}^p$. (As a convenient location normalisation, we set $f_0(0) = 0$.) This model evidently nests (2.1), by taking $f_0(z_t) = \Phi_0 z_t$ and $\boldsymbol{f}_1(\boldsymbol{z}_{t-1}) = c + \sum_{i=1}^k \Phi_i z_{t-i}$. Another important special case arises when $\boldsymbol{f}_1(\boldsymbol{z}_{t-1}) = \sum_{i=1}^k f_i(z_{t-i})$ is additively time-separable, as considered in Duffy and Mavroeidis (2024). But while this separability facilitates an extension of the Granger–Johansen representation theorem to these nonlinear SVARs, it is not necessary for the results that follow. The only separability that we require here is between z_t, $\boldsymbol{z}_{t-1}$ and ε_t.

We develop the following running example throughout the rest of the paper.

Example 2.1 (nonlinear Phillips curve)**.** The Phillips curve is a key component of any macroeconomic model. It provides a causal link between aggregate output and prices, and is thus essential in modelling the monetary policy transmission mechanism. Its name derives from the seminal contribution of Phillips (1958), who proposed the following simple *nonlinear* relationship between (wage) inflation, π_t^w, and labour market tightness as measured by the unemployment rate, u_t,

$$\pi_t^w = a + b\left(\frac{1}{u_t}\right)^c. \tag{2.3}$$

Several recent contributions have used the vacancy-to-unemployment ratio, $\theta_t := v_t/u_t$, as an alternative measure of tightness, and price (instead of wage) inflation, π_t, see e.g. Ball, Leigh, and Mishra (2022), Benigno and Eggertsson (2023), and Beaudry et al. (2025). These papers employ alternative functional forms for (2.3), and introduce additional dynamics, inflation expectations, and other shocks.[2]

Here we consider a stylised version of the piecewise linear Phillips curve proposed in Benigno and Eggertsson (2023),

$$\pi_t - \pi = \begin{cases} \kappa \log \theta_t + \eta_{\pi,t}, & \text{if } \theta_t \leq 1 \text{ ('normal')} \\ \kappa^{\text{tight}} \log \theta_t + \eta_{\pi,t}, & \text{if } \theta_t > 1 \text{ ('labour shortage')} \end{cases} \tag{2.4}$$

where π denotes steady state or target inflation, and $\eta_{\pi,t}$ an exogenous shock. Despite differences in specification, the fundamental identification problem in all such models remains the same. Insofar as inflation and tightness may plausibly be determined *simultaneously*, the r.h.s. of (2.4) cannot (in general) be identified as though it were a (nonlinear) regression. Simultaneous causation can instead be addressed by incorporating both (2.4), and the corresponding reverse (causal) model of the effect of inflation on tightness, into a nonlinear function $f_0(z_t)$, where $z_t = (\log \theta_t, \pi_t)^\mathsf{T}$, yielding our specification for the l.h.s. of (2.2). □

As noted in the introduction, we term (2.2) an *endogneously* nonlinear SVAR, because of the possible nonlinearity on the l.h.s. of the model, i.e. in the endogenous variables z_t. Were the model

[2]Ball et al. (2022) use a third order polynomial in $\log \theta_t$, Benigno and Eggertsson (2023) a piecewise linear function in $\log \theta_t$ with a kink at $\theta_t = 1$, and Beaudry et al. (2025) consider both these specifications.

instead required to be linear in the endogenous variables, so that $f_0(z) = \Phi_0 z$, identification of the model parameters would be as straightforward as it is in the linear SVAR; and along the lines of Remark 2.1(ii) above, the assumption that $\{\varepsilon_t\}$ is independent across time could be weakened to one of $\{\varepsilon_t\}$ being merely a martingale difference sequence (with respect to the filtration generated by $\{z_t\}$). However, in imposing linearity on the l.h.s., we would lose the possibilities for endogenous regime switching, asymmetric impact multipliers, and of handling occasionally binding constraints, which the general model (2.2) affords. We accordingly want to permit f_0 to be nonlinear: a consequence of which is that that *independence* across time of $\{\varepsilon_t\}$ becomes necessary for the parameters of (2.2) to be identified.

We thus continue to maintain that the structural shocks $\{\varepsilon_t\}$ are i.i.d. with $\mathbb{E}\varepsilon_t = 0$ and $\mathbb{E}\varepsilon_t\varepsilon_t^{\mathsf{T}} = I_p$, and a (Lebesgue) density $\varrho \in \mathscr{R}$. The parameter space for the model (2.2) then consists of collections $\mathscr{F}_0 \ni f_0$ and $\boldsymbol{\mathscr{F}}_1 \ni \boldsymbol{f}_1$ of functions $\mathbb{R}^p \to \mathbb{R}^p$ and $\mathbb{R}^{kp} \to \mathbb{R}^p$, and a collection of densities $\mathscr{R} \ni \varrho$ supported on $\mathbb{R}^p$, which under our regularity conditions, together determine the conditional density

$$\varphi_{z_t|\boldsymbol{z}_{t-1}}(\xi \mid \boldsymbol{\xi}_{-1}) = \varrho[f_0(\xi) - \boldsymbol{f}_1(\boldsymbol{\xi}_{-1})] \cdot |\det Df_0(\xi)|. \tag{2.5}$$

We continue to regard two alternative parametrisations of the model as being observationally equivalent if they yield the same conditional density.

Our regularity conditions on the model parameter space, which are sufficient to ensure that the conditional density (2.5) exists (and is unique up to the usual a.e. equivalence), are as follows.

Assumption PS. *$\mathscr{F}_0$, $\boldsymbol{\mathscr{F}}_1$ and $\mathscr{R}$ collect every function such that:*

1. *$\tilde{f}_0 \in \mathscr{F}_0$ and $\tilde{\boldsymbol{f}}_1 \in \boldsymbol{\mathscr{F}}_1$ are locally Lipschitz (continuous);*
2. *$\tilde{f}_0 \in \mathscr{F}_0$ is a bijection $\mathbb{R}^p \to \mathbb{R}^p$, $\tilde{f}_0(0) = 0$, and $\det D\tilde{f}_0(z) \neq 0$ for almost every $z \in \mathbb{R}^p$;*
3. *$\tilde{\varrho} \in \mathscr{R}$ is continuously differentiable, with $\tilde{\varrho}(\varepsilon) > 0$ for all $\varepsilon \in \mathbb{R}^p$, and*

$$\int_{\mathbb{R}^p} \tilde{\varrho}(\varepsilon)\,\mathrm{d}\varepsilon = 1, \qquad \int_{\mathbb{R}^p} \varepsilon\tilde{\varrho}(\varepsilon)\,\mathrm{d}\varepsilon = 0, \qquad \int_{\mathbb{R}^p} \varepsilon\varepsilon^{\mathsf{T}}\tilde{\varrho}(\varepsilon)\,\mathrm{d}\varepsilon = I_p.$$

Remark 2.2. (i). Local Lipschitzness implies that $\tilde{f}_0$ and $\tilde{\boldsymbol{f}}_1$ are differentiable almost everywhere (a.e.). The r.h.s. of (2.5) is therefore defined (when $f_0 = \tilde{f}_0$ and $\boldsymbol{f}_1 = \tilde{\boldsymbol{f}}_1$) at least a.e., which is sufficient to pin down the conditional density $\varphi_{z_t|\boldsymbol{z}_{t-1}}$, since the latter is itself only uniquely defined up to an a.e. equivalence. (See Appendix A for further details.) Our smoothness conditions and support conditions on the density $\tilde{\varrho}$ (which accord with those of Matzkin, 2008) are maintained only for convenience, and could very likely also be relaxed in this same direction.

(ii). Since the nonlinear SVAR (2.2) is a (dynamic) nonlinear SEM, our work relates closely to the literature on identification in such models: particularly Matzkin (2008, 2015) and Berry and Haile (2018). Here we have deliberately relaxed the assumption that the functions $\tilde{f}_0$ and $\tilde{\boldsymbol{f}}_1$ are (at least once) continuously differentiable, which is standard in that literature, to allow our results to accommodate models that are continuous but merely piecewise differentiable, such as the piecewise affine SVARs introduced in Section 3 below.

(iii). We naturally require $\tilde{f}_0$ to be invertble, which ensures that the model always yields a solution for the endogenous variables z_t, irrespective of the values of the predetermined variables $\boldsymbol{z}_{t-1}$ and the structural shocks ε_t. Requiring $\det D\tilde{f}_0(z) \neq 0$ (a.e.) merely excludes certain irregular cases (our assumptions also imply that this quantity must have the same sign a.e.).

Regarding the parameters $(f_0, \boldsymbol{f}_1, \varrho)$ that generated $\{z_t\}$ in (2.2), as distinct from the *entirety* of the model parameter space, we also maintain the following.

Assumption DGP. *$(f_0, \boldsymbol{f}_1, \varrho)$ are such that:*

1. *$\boldsymbol{f}_1 : \mathbb{R}^{kp} \to \mathbb{R}^p$ is surjective, with $\operatorname{rk} D\boldsymbol{f}_1(\boldsymbol{z}) = p$ for almost every $\boldsymbol{z} \in \mathbb{R}^{kp}$;*
2. *f_0^{-1} is locally Lipschitz; and*
3. *ϱ has a* local *maximum at some $\varepsilon^* \in \mathbb{R}^p$, and is twice continuously differentiable in a neighbourhood of ε^*, with negative definite Hessian there.*

Remark 2.3. (i). We interpret DGP.1 as requiring that there be sufficient dependence of the r.h.s. of the model (i.e. on the conditional mean of $f_0(z_t)$) on the predetermined variables $\boldsymbol{z}_{t-1}$, in both a 'global' and 'local' sense. (Note that this is only a requirement on the $\boldsymbol{f}_1$ that actually generated the data, which need *not* be satisfied by all members of $\boldsymbol{\mathscr{F}}_1$). For a simple illustration of why some such condition cannot be avoided, consider an extreme case in which $\boldsymbol{f}_1(\boldsymbol{z}) = 0$ for all $\boldsymbol{z} \in \mathbb{R}^{kp}$, so that the r.h.s. of (2.2) does not depend on $\boldsymbol{z}_{t-1}$ at all. Then because $z_t = f_0^{-1}(\varepsilon_t)$ will be i.i.d. and independent of $\boldsymbol{z}_{t-1}$, so too will be

$$\tilde{\varepsilon}_t \coloneqq \tilde{f}_0(z_t) = \tilde{f}_0[f_0^{-1}(\varepsilon_t)]$$

for *every* $\tilde{f}_0 \in \mathscr{F}_0$. Beyond requiring $\tilde{f}_0$ to be scale- and location-normalised such that $\mathbb{E}\tilde{\varepsilon}_t = 0$ and $\mathbb{E}\tilde{\varepsilon}_t\tilde{\varepsilon}_t^{\mathsf{T}} = I_p$, the model would therefore yield no meaningful identifying restrictions on $\tilde{f}_0$.

(ii). DGP.2 is a weak regularity condition on the inverse of f_0, which would e.g. be automatically satisfied if f_0 were continuously differentiable with $\det Df_0(z) \neq 0$ for all $z \in \mathbb{R}^p$.

(iii). DGP.3 would clearly be satisfied if ε_t were Gaussian; but note that only a well-behaved *local* maximum is required for this condition to hold. The main purpose of this assumption is to allow us to deduce that $u = u^*$ from merely the equality $f_U(u) = f_U(u^*)$, and further regulate the behaviour of f_U in the vicinity of u^*. Though their model and proofs differ significantly from ours, it is noteworthy that a similar assumption is introduced by Berry and Haile (2018) as their Condition M (see also their Corollary 2).

Remarkably, despite the far greater flexibility afforded by the nonlinear parametrisation of (2.2), under the foregoing conditions we obtain the following, effectively identical characterisation of observational equivalence, to that of the linear SVAR (2.1); the proof of which appears in Appendix A.

Theorem 2.2. *Suppose* PS *and* DGP *hold, and let $\tilde{f}_0 \in \mathscr{F}_0$ and $\tilde{\boldsymbol{f}}_1 \in \boldsymbol{\mathscr{F}}_1$. Then there exists a $\tilde{\varrho} \in \mathscr{R}$ such that $(\tilde{f}_0, \tilde{\boldsymbol{f}}_1, \tilde{\varrho})$ is observationally equivalent to $(f_0, \boldsymbol{f}_1, \varrho)$, if and only if there exists a $Q \in \mathbb{O}(p)$ such that*

$$\tilde{f}_0(z) = Qf_0(z), \ \forall z \in \mathbb{R}^p, \qquad\qquad \boldsymbol{f}_1(\boldsymbol{z}) = Q\tilde{\boldsymbol{f}}_1(\boldsymbol{z}), \ \forall \boldsymbol{z} \in \mathbb{R}^{kp}. \tag{2.6}$$

Remark 2.4. (i). Here we are asking whether for given candidate functions $(\tilde{f}_0, \tilde{\boldsymbol{f}}_1)$, it is possible to find a distribution $\tilde{\varrho} \in \mathscr{R}$ for the structural shocks such that

$$\varrho[f_0(\xi) - \boldsymbol{f}_1(\boldsymbol{\xi}_{-1})] \cdot |\det Df_0(\xi)| = \tilde{\varrho}[\tilde{f}_0(\xi) - \tilde{\boldsymbol{f}}_1(\boldsymbol{\xi}_{-1})] \cdot |\det D\tilde{f}_0(\xi)|$$

holds for almost every $\xi \in \mathbb{R}^p$ and $\boldsymbol{\xi}_{-1} \in \mathbb{R}^{kp}$ (cf. Theorem 2.1 above). The $\tilde{\varrho}$ delivering this equivalence will, for Q as in (2.6), be given by the density of

$$\tilde{\varepsilon}_t = \tilde{f}_0(z_t) - \tilde{\boldsymbol{f}}_1(\boldsymbol{z}_{t-1}) = Q[f_0(z_t) - \boldsymbol{f}_1(\boldsymbol{z}_{t-1})] = Q\varepsilon_t,$$

which under our assumptions will also lie in $\mathscr{R}$. This implies that the introduction of further (e.g. parametric) assumptions on the set $\mathscr{R}$ of allowable densities would not yield any further tightening of our characterisation of observational equivalence, provided that $\mathscr{R}$ remains closed under orthogonal transformations of the variables: as would e.g. be the case if $\mathscr{R}$ were restricted to the set of Gaussian densities on $\mathbb{R}^p$ (with mean zero and identity covariance).

(ii). The foregoing is a nonparametric identification result, in the sense that neither $(f_0, \boldsymbol{f}_1)$, nor the distribution ϱ of the shocks, are assumed to have any particular (known) parametric form. In practice, however, we would expect the model (2.2) to be formulated parametrically, if only because the limited length of the time series available, for most macroeconomic applications, render genuine nonparametric estimation infeasible. In the abstract setting of Theorem 2.2, these parametric functional form and/or distributional assumptions can be understood as restrictions on the sets $\mathscr{F}_0$, $\boldsymbol{\mathscr{F}}_1$ and $\mathscr{R}$. The conclusion of the theorem continues to hold in such cases, provided that $\mathscr{R}$ is not so (unusually) constrained that it fails to satisfy the invariance condition noted in the previous remark. See Section 3 below for the discussion of a class of parametric models (for f_0 and $\boldsymbol{f}_1$) for which the conditions required by the theorem may be verified straightforwardly.

(iii). As noted above, a consequence of the Markov property of the SVAR is that the appropriate notion of 'observational equivalence' in our setting refers only to the distribution $z_t \mid \boldsymbol{z}_{t-1}$ of the endogenous variables conditional on the exogenous variables; it therefore coincides exactly with that employed by Matzkin (2008) in the context of a (non-dynamic) nonlinear SEM. This allows the proof of Theorem 2.2 to be approached in the same manner as that of establishing identification in nonlinear SEMs, a connection that we draw out more fully in Appendix A. Relative to the results in the existing literature, we are able to obtain a much tighter characterisation of observational equivalence because of the separability between z_t and $\boldsymbol{z}_{t-1}$.

(iv). Should (2.6) fail to hold, then there will be at least some realisations of $\{z_t\}$ for which the likelihoods of $(\tilde{f}_0, \tilde{\boldsymbol{f}}_1, \tilde{\varrho})$ and $(f_0, \boldsymbol{f}_1, \varrho)$ will be distinct, and so the data is to this extent informative about these two alternative parametrisations of the model. However, we would not claim, on the basis of this theorem *alone*, that the parameters of the model are consistently estimable up to an orthogonal transformation. While it seems reasonable to suppose that consistent nonparametric estimation of the model would be possible (under suitable regularity conditions) when $\{z_t\}$ is stationary and ergodic, the familiar connection between identification and consistent estimation is attenuated when $\{z_t\}$ possesses stochastic (or indeed, deterministic) trends, because of the non-recurrence of those trends in higher dimensions (see Bingham, 2001, Sec. 6;

Gao and Phillips, 2013, p. 62). Consistent estimation of the model parameters (up to Q) would in such cases likely require further restrictions on $(f_0, \boldsymbol{f}_1)$, such as those sufficient to ensure that $\{z_t\}$ is indeed stationary and ergodic (for a discussion of such conditions in this context, see Duffy, Mavroeidis, and Wycherley, 2023, and the references cited therein).

(v). While the shocks are assumed to have constant variance (equal to I_p), the nonlinearity on the l.h.s. of the model allows certain forms of regime-dependent volatility to be accommodated. Theorem 2.2 may also be extended, with relatively minor modifications, to cover the case in which ε_t is replaced by $\sigma(v_{t-1})\varepsilon_t$, where v_{t-1} is an observed, predetermined process (possibly but not necessarily formed from elements of $\boldsymbol{z}_{t-1}$), and $\sigma(v)$ is a positive diagonal matrix for every $v \in \mathbb{R}^{d_v}$.

2.3 Orthogonal reduced-form parametrisation

Analogously (though not identically) to the 'orthogonal reduced-form parametrisation' (Arias et al., 2018, Sec. 2.3) of the linear SVAR, Theorem 2.2 suggests the following convenient reparametrisation of the endogenously nonlinear SVAR. Let $z_0 \in \mathbb{R}^p$ be fixed, and a point at which f_0 is (assumed to be) differentiable, with full rank Jacobian $Df(z_0)$. By the QR decomposition, we have $Df(z_0) = Q^{\mathsf{T}}L$, where L is lower triangular, and $Q \in \mathbb{O}(p)$; define $(g_0, \boldsymbol{g}_1) \coloneqq (Qf_0, Q\boldsymbol{f}_1)$. Multiplying (2.2) through by Q, we may reformulate the model as

$$g_0(z_t) = \boldsymbol{g}_1(z_{t-1}) + Q\varepsilon_t \tag{2.7}$$

where now g_0 is restricted such that $Dg(z_0)$ is lower triangular (which need hold *only* at that chosen z_0), and $Q \in \mathbb{O}(p)$.

This yields an equivalent parametrisation of the model, in which the parameter spaces for $\boldsymbol{g}_1 \in \boldsymbol{\mathscr{F}}_1$ and $\varrho \in \mathscr{R}_1$ remain as before, but now $\mathscr{F}_0$ is additionally restricted (beyond PS.1–2) to functions $g_0 : \mathbb{R}^p \to \mathbb{R}^p$ for which $Dg_0(z_0)$ is lower triangular (at the nominated $z_0 \in \mathbb{R}^p$); let $\mathscr{F}_0^{(z_0)}$ denote the resulting parameter space for g_0. To exactly offset this restriction, we now have the additional parameter $Q \in \mathbb{O}(p)$, so that we may equivalently regard the nonlinear SVAR as being parametrised by $(g_0, \boldsymbol{g}_1, Q, \varrho) \in \mathscr{F}_0^{(z_0)} \times \boldsymbol{\mathscr{F}}_1 \times \mathbb{O}(p) \times \mathscr{R}$. The import of Theorem 2.2 here is that the parameters $(g_0, \boldsymbol{g}_1) \in \mathscr{F}_0^{(z_0)} \times \boldsymbol{\mathscr{F}}_1$ are *exactly* identified by data on $\{z_t\}$, with the non-identified part of the structural parameters being transferred entirely to Q. The 'nonlinear SVAR identification problem' can thus be framed precisely as one of finding sufficient restrictions to pin down Q, from which the structural parameters may then be recovered, via $(f_0, \boldsymbol{f}_1) = (Q^{\mathsf{T}}g_0, Q^{\mathsf{T}}\boldsymbol{g}_1)$.

The reparametrisation (2.7) provides a convenient perspective from which to import various approaches to identifying Q from the linear SVAR literature. For the most part, these apply directly to the present setting, with little modification required. The following example illustrates how it remains possible to identify impulse responses via external instruments, without requiring any additional assumptions relative to those needed to identify the linear VAR.

Example 2.2 (external instruments)**.** Suppose that w_t is a (scalar) 'external instrument': an observed (stationary) process that is assumed to be contemporaneously correlated with the first structural shock, but not with any of the others (see e.g. Stock and Watson, 2018, p. 931).

Defining

$$u_t \coloneqq g_0(z_t) - \boldsymbol{g}_1(z_{t-1}) = Q\varepsilon_t \tag{2.8}$$

which is identified, in view of (2.7), we must have

$$\delta \coloneqq \mathbb{E}u_t w_t = Q\mathbb{E}\varepsilon_t w_t = Qe_1\alpha = q_1\alpha,$$

where q_1 denotes the first column of Q, and $\alpha = \mathbb{E}\varepsilon_{1t}w_t \neq 0$. Since $\delta = \mathbb{E}u_t w_t$ is identified, so too is $q_1 = \delta/\|\delta\|$, and we can further recover $\varepsilon_{1t} = q_1^{\mathsf{T}}u_t$.

Since the distribution of u_t in (2.8) is identified, so too is the conditional distribution

$$u_t \mid \{\varepsilon_{1t} = \bar{\varepsilon}_1\} =_d u_t \mid \{q_1^{\mathsf{T}}u_t = \bar{\varepsilon}_1\}.$$

For given values of $\boldsymbol{z}_{t-1} = \bar{\boldsymbol{z}}$ and $\bar{\varepsilon}_1$, the distribution of the counterfactual quantity

$$z_t(\bar{\boldsymbol{z}}, \bar{\varepsilon}_1) \mid \{\boldsymbol{z}_{t-1} = \bar{\boldsymbol{z}}, \varepsilon_{1t} = \bar{\varepsilon}_1\} =_d g_0^{-1}(\boldsymbol{g}_1(\bar{\boldsymbol{z}}) + u_t) \mid \{q_1^{\mathsf{T}}u_t = \bar{\varepsilon}_1\}$$

depends only on g_0, $\boldsymbol{g}_1$ and the distribution of $u_t \mid \{q_1^{\mathsf{T}}u_t = \bar{\varepsilon}_1\}$, all of which are identified. In this way, the impact multipliers of ε_{1t} may be recovered; the impulse responses at further horizons depend, by the Markov property, additionally only on the conditional distribution of $z_t \mid \boldsymbol{z}_{t-1}$, which is trivially identified. □

3 Piecewise affine SVARs

3.1 Endogenous regime switching

Here we introduce a class of endogenously regime-switching models, in which the conditions required for our results may be verified relatively straightforwardly. Models of this form have been used recently to study monetary policy under an occasionally binding constraint on nominal interest rates: see Mavroeidis (2021), Aruoba et al. (2022) and Carriero, Clark, Marcellino, and Mertens (2025).

Suppose now that the l.h.s. of the nonlinear SVAR

$$f_0(z_t) = \boldsymbol{f}_1(\boldsymbol{z}_{t-1}) + \varepsilon_t \tag{3.1}$$

is specified as

$$f_0(z) = \sum_{\ell=1}^{L} \mathbf{1}\{z \in \mathscr{Z}^{(\ell)}\}(\bar{\phi}_0^{(\ell)} + \Phi_0^{(\ell)}z), \tag{3.2}$$

for $\{\mathscr{Z}^{(\ell)}\}_{\ell=1}^L$ a collection of convex sets that partition $\mathbb{R}^p$, $\{\bar{\phi}_0^{(\ell)}\}_{\ell=1}^L \subset \mathbb{R}^p$ and $\{\Phi_0^{(\ell)}\}_{\ell=1}^L \subset \mathbb{R}^{p\times p}$. When these parameters are such that f_0 is continuous, we shall say that f_0 is a *piecewise affine* function. (We do not consider cases in which f_0 may be discontinuous, so continuity should always be taken as implied.) The model may then be regarded as consisting of L 'regimes' demarcated by the sets $\{\mathscr{Z}^{(\ell)}\}_{\ell=1}^L$. Which of those regimes is operative in period t, i.e. the value of $\ell_t \in \{1, \ldots, L\}$ such that

$$f_0(z_t) = \bar{\phi}_i^{(\ell_t)} + \Phi_i^{(\ell_t)}z_t$$

is determined *jointly* with the value of z_t. For this reason, we say that there is *endogenous* switching between the L regimes, as distinct from the *exogenous* regime switching that would result if ℓ_t were determined prior to the realisation of z_t. The situation here is thus markedly different from the regime-switching SVARs considered in the previous literature, which as noted in the introduction, can generally be written in the form

$$\Phi_0(s_{t-1})z_t = c(s_{t-1}) + \sum_{i=1}^{k} \Phi_i(s_{t-1})z_t + \varepsilon_t,$$

where s_{t-1} is determined prior to ε_t and z_t (see e.g. Auerbach and Gorodnichenko, 2012; Caggiano, Castelnuovo, Colombo, and Nodari, 2015; Bruns and Piffer, 2024).

Example 2.1 (nonlinear Phillips curve; ctd)**.** The nonlinear Phillips curve of Benigno and Eggertsson (2023) displayed in (2.4) is piecewise affine and continuous at the kink $\theta_t = 1$, which the authors refer to as the 'Beveridge threshold'. Their model thus delineates two distinct labour market regimes: a 'normal' regime ($\ell_t = 1$), when the labour market is slack, $\log \theta_t \leq 0$, and a 'labour shortage' regime ($\ell_t = 2$) in which $\log \theta_t > 1$. The regime ℓ_t is entirely driven by the *contemporaneous* value of the endogenous variable $\log \theta_t$, and so the regime-switching is genuinely endogenous. Their Phillips curve (2.4) can also be written as

$$\pi_t = \pi + \kappa^{(\ell_t)} \log \theta_t + \eta_t. \tag{3.3}$$

where $\kappa^{(1)} = \kappa$ and $\kappa^{(2)} = \kappa^{\text{tight}}$. Contrast this with an alternative specification in which the slope of the Phillips curve is determined by *past* values of $\log \theta_t$, e.g.

$$\pi_t = \pi + \kappa^{(\ell_{t-1})} \log \theta_t + \eta_t. \tag{3.4}$$

Conditional on the past (i.e. on time $t-1$), (3.4) is linear in $z_t = (\log \theta_t, \pi_t)^{\mathsf{T}}$, and so shocks to $\log \theta_t$ will have the same proportional effect $\kappa^{(\ell_{t-1})}$ on π_t, irrespective of their sign; whereas in (3.3) the impact of the shocks will vary additionally (and nonlinearly) depending on the initial (i.e. pre-shock) proximity of tightness to the Beveridge threshold. □

3.2 Identification

We would like primitive conditions that ensure f_0 in (3.2) satisfies the requirements PS.1-2 and DGP.2 of Theorem 2.2: namely, that both it and its inverse should be locally Lipschitz, and that it should be (globally) invertible, with $\det Df_0(z) \neq 0$ a.e. Two important special cases of (3.2), for which these conditions may be readily verified, are those of:

- a (continuous) *piecewise linear* function, in which there exists a basis $\{a_i\}_{i=1}^p$ for $\mathbb{R}^p$ such that each $\mathscr{Z}^{(\ell)}$ can be written as a union of cones of the form

$$\mathscr{C}_{\mathcal{I}} \coloneqq \{z \in \mathbb{R}^p \mid a_i^{\mathsf{T}} z \geq 0,\ \forall i \in \mathcal{I} \text{ and } a_i^{\mathsf{T}} z < 0,\ \forall i \notin \mathcal{I}\} \tag{3.5}$$

  where $\mathcal{I}$ ranges over the subsets of $\{1, \ldots, p\}$, and $\bar{\phi}_i^{(\ell)} = 0$ for all i and ℓ; and

- a (continuous) *threshold affine* function, in which there exists an $a \in \mathbb{R}^p \backslash \{0\}$ and thresholds $\{\tau_\ell\}_{\ell=0}^L$ with $\tau_\ell < \tau_{\ell+1}$, $\tau_0 = -\infty$ and $\tau_L = +\infty$, such that

$$\mathscr{Z}^{(\ell)} = \{z \in \mathbb{R}^p \mid a^\mathsf{T} z \in (\tau_{\ell-1}, \tau_\ell]\},$$

  i.e. the sets $\{\mathscr{Z}^{(\ell)}\}$ take the forms of 'bands' in $\mathbb{R}^p$. (In typical examples, $a = e_{p,i}$, i.e. it picks out one 'threshold variable' from the elements of z_t.)

Because the boundaries between the regimes are then linear subspaces, ensuring the continuity of f_0 is a straightforward matter of linearly restricting the elements of $\{\bar{\phi}_0^{(\ell)}\}_{\ell=1}^L$ and $\{\Phi_0^{(\ell)}\}_{\ell=1}^L$ such that the values prescribed by adjacent regimes agree on those boundaries; see the next appearance of Example 2.1 for an illustration. Regarding our remaining requirements on f_0, for these it is necessary and sufficient that

$$\operatorname{sgn}\det \Phi_0^{(\ell)} = \operatorname{sgn}\det \Phi_0^{(1)} \neq 0, \ \forall \ell \in \{1, \dots, L\}. \tag{3.6}$$

See Proposition 3.1 below; we note that equivalence of the preceding with the invertibility of f_0 follows directly from Theorems 1 and 4 of Gouriéroux, Laffont, and Monfort (1980), and that since $Df_0(z) = \sum_{\ell=1}^L \mathbf{1}\{z \in \mathscr{Z}^{(\ell)}\}\Phi_0^{(\ell)}$ a.e., the Jacobian is clearly invertible a.e.

Example 2.1 (nonlinear Phillips curve; ctd)**.** The nonlinear Phillips curve (2.4) is piecewise linear, with $z_t = (\log\theta_t, \pi_t)^\mathsf{T}$ and two regimes determined by (3.5) with $a = e_1 = (1,0)^\mathsf{T}$. (2.4) only specifies the first element of the bivariate map $f_0(z_t)$. If the second element is also modelled as piecewise linear, with the regimes also determined by the sign of $a^\mathsf{T} z_t$, then f_0 admits the representation (3.2). To ensure continuity at the regime boundary where $a^\mathsf{T} z_t = \log\theta_t = 0$, we need the equality

$$\bar{\phi}_0^{(1)} + \begin{bmatrix}\Phi_{0,1}^{(1)} & \Phi_{0,2}^{(1)}\end{bmatrix}\begin{bmatrix}0\\ \pi_t\end{bmatrix} = \bar{\phi}_0^{(2)} + \begin{bmatrix}\Phi_{0,1}^{(2)} & \Phi_{0,2}^{(2)}\end{bmatrix}\begin{bmatrix}0\\ \pi_t\end{bmatrix}$$

to hold for all values of $\pi_t \in \mathbb{R}$, where $\Phi_0^{(\ell)} = [\Phi_{0,1}^{(\ell)}, \Phi_{0,2}^{(\ell)}]$. This entails

$$\bar{\phi}_0^{(1)} - \bar{\phi}_0^{(2)} = 0 \qquad\qquad \Phi_{0,2}^{(1)} - \Phi_{0,2}^{(2)} = 0,$$

and we may also impose $\bar{\phi}_0^{(1)} = 0$, for the location normalisation $f_0(0) = 0$. Or in other words, only the coefficients on the regime-determining variable $\log\theta_t$ may change at the threshold, leading to the (non-redundant) specification

$$\Phi_0^{(\ell)} = [\Phi_{0,1}^{(\ell)}, \Phi_{0,2}], \ \ell \in \{1,2\} \tag{3.7}$$

in which the second column of coefficients is regime-invariant. □

The SVAR specification (3.1)–(3.2) thus provides a flexible but tractable means of introducing nonlinearity into an SVAR model. This is particularly the case if we also specify that the r.h.s. should be additively time-separable, and of the same functional form as the l.h.s., so that

$$f_0(z_t) = \boldsymbol{f}_1(\boldsymbol{z}_{t-1}) + \varepsilon_t = c + \sum_{i=1}^k f_i(z_{t-i}) + \varepsilon_t \tag{3.8}$$

where now, for *every* $i \in \{0, \dots, k\}$,

$$f_i(z) = \sum_{\ell=1}^{L} \mathbf{1}\{z \in \mathscr{Z}^{(\ell)}\}(\bar{\phi}_i^{(\ell)} + \Phi_i^{(\ell)} z), \tag{3.9}$$

is (continuous) piecewise affine. (Note that there is no need to additionally index the regimes $\mathscr{Z}^{(\ell)}$ by i here, since if the partitions $\{\mathscr{Z}_i^{(\ell)}\}_{\ell=1}^{L_i}$ did vary across i, we could always find a mutual refinement such that (3.9) held for all i.) We term this model a *piecewise affine SVAR*; with *piecewise linear* and *threshold affine SVARs* corresponding to those cases where the f_i's are either all piecewise linear or all threshold affine functions, respectively.

The conditions PS.1 and DGP.1 imposed by Theorem 2.2 on $\boldsymbol{f}_1(\boldsymbol{z}_{t-1}) = \sum_{i=1}^{k} f_i(z_{t-i})$ are rather less taxing than those imposed on f_0. Under the specification (3.9), continuity is readily imposed, and then automatically implies Lipschitz continuity. Moreover, $D\boldsymbol{f}_1(\boldsymbol{z}_{t-1})$ a.e. exists and satisfies

$$D\boldsymbol{f}_1(\boldsymbol{z}) = \begin{bmatrix} \Phi_1^{(\ell_1)} & \Phi_2^{(\ell_2)} & \cdots & \Phi_k^{(\ell_k)} \end{bmatrix}$$

for some $\ell_i \in \{1, \dots, L\}$ depending on $\boldsymbol{z}$, and so it is easy to verify whether $\operatorname{rk} D\boldsymbol{f}_1(\boldsymbol{z}) = p$ a.e. (or, since this holds generically, to test the null hypothesis of a deficient rank). In practice, this may be analysed more straightforwardly on the basis of the coefficients on the first lag or two of z_t, which may themselves be sufficient to satisfy this condition.

Verifying the (global) surjectivity condition on $\boldsymbol{f}_1$ is a little more challenging, because of the apparent absence of a counterpart to (3.6) for this case. In the special case of a model with only one lag, surjectivity of $z_{t-1} \mapsto f_1(z_{t-1})$ is equivalent to (3.6), but which though easy to check is far more than is necessary for surjectivity when additional lags are present. Alternatively, if some elements of z_t enter f_i linearly, as will often be the case in practice (as in our next example), then surjectivity holds so long as the coefficient vectors associated with (at least) p of these variables (drawn from across the k lags of z_t appearing on the r.h.s.) form a rank p matrix.

Example 3.1 (occasionally binding constraint)**.** Mavroeidis (2021) proposed the censored and kinked structural VAR (CKSVAR), to model the effects of the zero lower bound (ZLB) constraint on monetary policy: see also Aruoba et al. (2022) and Carriero et al. (2025). In his setting, y_t is a scalar variable whose positive part $y_t^+ \coloneqq \max\{y_t, 0\}$ coincides with the central bank's policy rate (constrained to be non-negative), while its (latent) negative part $y_t^- \coloneqq \min\{y_t, 0\}$ is the 'shadow rate', which summarises the stance of monetary policy desired by the central bank when the ZLB binds, to be engineered via 'unconventional' policy, such as asset purchases. The remaining variables in the model are collected in the $(p-1)$-dimensional vector x_t, in his case the inflation and unemployment rates; we then set $z_t = (y_t, x_t^{\mathsf{T}})^{\mathsf{T}}$.

To allow for possibility that the ZLB might actually constrain monetary policy, y_t^+ and y_t^- are permitted to enter the model with different coefficients (in possibly all p equations),

$$\phi_0^+ y_t^+ + \phi_0^- y_t^- + \Phi_0^x x_t = c + \sum_{i=1}^{k} [\phi_i^+ y_{t-i}^+ + \phi_i^- y_{t-i}^- + \Phi_i^x x_{t-i}] + u_t \tag{3.10}$$

where $\phi_i^{\pm} \in \mathbb{R}^p$ and $\Phi_i^x \in \mathbb{R}^{p \times (p-1)}$, for $i \in \{0, \dots, k\}$. This may be rendered as an instance of a

threshold affine SVAR by defining

$$\mathscr{Z}^{(1)} \coloneqq \mathscr{Z}^{-} = \{(y,x) \in \mathbb{R}^p \mid y \leq \tau_1\} \qquad \Phi_i^{(1)} \coloneqq [\phi_i^{-}, \Phi_i^x] \tag{3.11a}$$

$$\mathscr{Z}^{(2)} \coloneqq \mathscr{Z}^{+} = \{(y,x) \in \mathbb{R}^p \mid y > \tau_1\} \qquad \Phi_i^{(2)} \coloneqq [\phi_i^{+}, \Phi_i^x], \tag{3.11b}$$

with $\tau_1 = 0$, and then setting $f_i(z) = \sum_{\ell=1}^{2} \mathbf{1}\{z \in \mathscr{Z}^{(\ell)}\}\Phi_i^{(\ell)} z$. (Because there are only two regimes, it may also be equivalently cast as a piecewise linear SVAR.) Here continuity of each f_i is guaranteed by the fact that $\Phi_i^{(1)}$ and $\Phi_i^{(2)}$ only differ by their first column; or equivalently by the linear restrictions $(\Phi_i^{(1)} - \Phi_i^{(2)})E_{-1} = 0$, for E_{-1} the final $p-1$ columns of I_p.

In Mavroeidis (2021), identification of the parameters of this model are complicated by the fact that y_t is only observed when $y_t > 0$; it is in effect censored at zero. His results therefore do not fall within the framework of Theorem 2.2, which implicitly assumes that z_t and $\boldsymbol{z}_{t-1}$ are observed on the entirety of their supports. However, the model (3.10)–(3.11) may (of course) also be applied to settings in which y_t is observed on both sides of the threshold τ_1, which may be treated as an additional unknown parameter to be identified and estimated. From the foregoing discussion, for f_0 to satisfy the conditions of Theorem 2.2, we would need only to verify that $\det \Phi_0^{(1)}$ and $\det \Phi_0^{(2)}$ are both nonzero, and have the same sign. Regarding $\boldsymbol{f}_1$: if $k \geq 2$ then it is sufficient to check (or rather, test) whether the $p \times k(p-1)$ matrix $[\Phi_1^x, \ldots, \Phi_k^x]$, formed from the coefficients on the lags of x_t, has rank p; whereas if $k = 1$, then we would need $\{\Phi_1^{(\ell)}\}$ to satisfy the same determinantal condition as $\{\Phi_0^{(\ell)}\}$ (a condition also sufficient when $k \geq 2$). □

3.3 Smooth transitions

The piecewise affine SVAR (3.8)–(3.9) may be extended to allow for 'smooth transitions' between the L regimes. In the literature on smooth transition (vector) autoregressive models, the conventional approach (e.g. Hubrich and Teräsvirta, 2013, Sec. 3.3) is to replace the indicator functions $\mathbf{1}\{z \in \mathscr{Z}^{(\ell)}\}$ by smooth maps $\Lambda^{(\ell)}(z)$, so that now

$$f_i^{\mathrm{ST}}(z) = \sum_{\ell=1}^{L} \Lambda^{(\ell)}(z)(\bar{\phi}_i^{(\ell)} + \Phi_i^{(\ell)} z),$$

where $\Lambda^{(\ell)}(z) \in [0,1]$ and $\sum_{\ell=1}^{L} \Lambda^{(\ell)}(z) = 1$ for all $z \in \mathbb{R}^p$, so that $f_i^{\mathrm{ST}}(z)$ is a always a smooth, convex combination of the affine functions $z \mapsto \bar{\phi}_i^{(\ell)} + \Phi_i^{(\ell)} z$, for $\ell \in \{1, \ldots, L\}$. However, the fact that the *gradient* of f_i^{ST} is *not* a convex combination of those underlying affine regimes makes it difficult to reduce the high-level conditions of Theorem 2.2 to a set of verifiable conditions on the underlying coefficient matrices, in the manner of (3.6). Indeed, as the simple example in Figure 3.1 illustrates, it may well be the case that f_0^{ST} is not invertible, even though its unsmoothed counterpart f_0 is.

As an alternative specification that allows for smooth transitions between regimes, but which also retains the simplicity – in terms of verifying the conditions for Theorem 2.2 – enjoyed by piecewise affine models, consider

$$f_{i,K}(z) \coloneqq \int_{\mathbb{R}^p} f_i(z+u)K(u)\,\mathrm{d}u \tag{3.12}$$

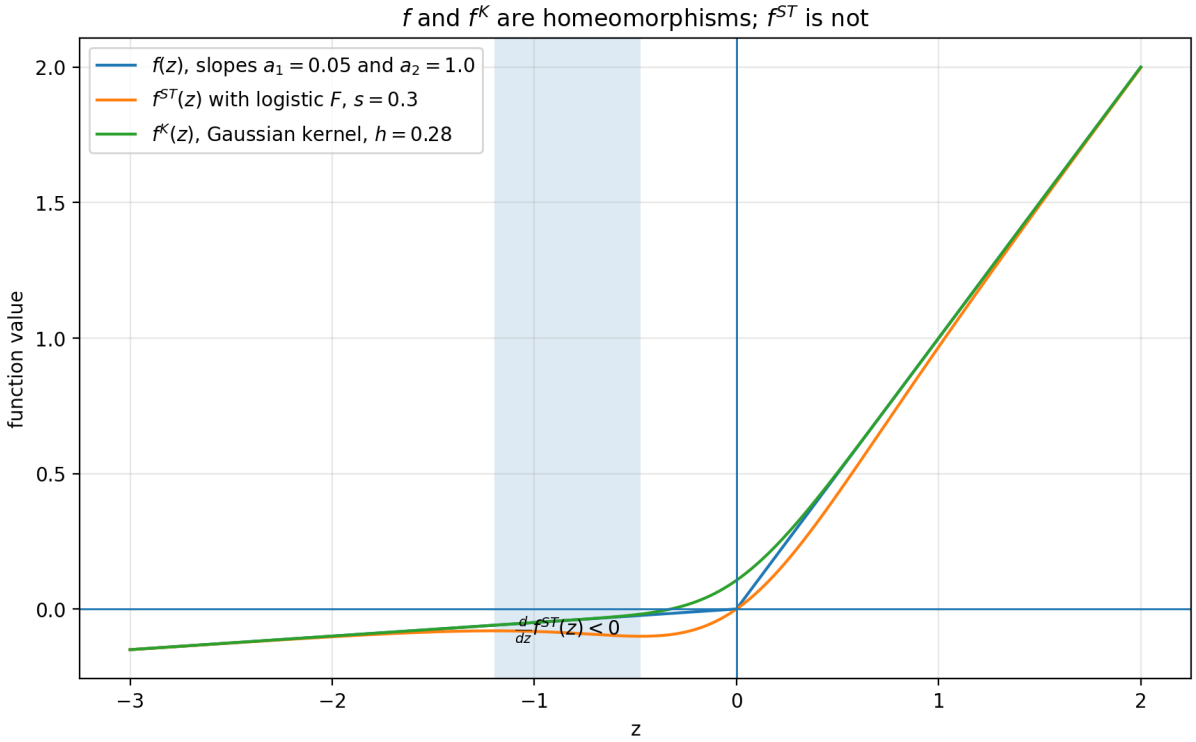


Figure 3.1: Plot of functions $f(z) = a_1 \min(z, 0) + a_2 \max(z, 0)$, $f^{ST}(z) = [1 - F(z)] a_1 z + F(z) a_2 z$, where $F(z) = \frac{1}{1+e^{z/s}}$ and $f^K(z) = \int f(z+u) K(u) du$, where $K(u) = h\varphi(u/h)$ and φ is the standard Gaussian pdf.

where f_i is a (continuous) piecewise affine function as in (3.9) above, and K is a smooth (kernel) density function with mean zero, with $m \geq 1$ continuous derivatives that satisfy the integrability condition

$$\int_{\mathbb{R}^p} \|u\| |\partial_{u_{\alpha_1}} \cdots \partial_{u_{\alpha_n}} K(u)| \, \mathrm{d}u < \infty, \tag{3.13}$$

where ∂_{u_i} denotes the partial derivative with respect to the ith element of $u \in \mathbb{R}^p$, for $\alpha_i \in \{1, \ldots, p\}$ and $1 \leq n \leq m$.

Our next result establishes that $f_{0,K}(z)$ is smooth (with as many continuous derivatives as K has), and moreover invertible if the determinantal condition (3.6) is satisfied; see Appendix B for the proof). Recall that a function is said to be bi-Lipschitz, if both it and its inverse are Lipschitz continuous.

Proposition 3.1. *Suppose that $f_0 : \mathbb{R}^p \to \mathbb{R}^p$ is as in (3.2), and is either a piecewise linear or threshold affine function. Then:*

(i) *f_0 is invertible and bi-Lipschitz if and only if (3.6) holds.*

Suppose that $K : \mathbb{R}^p \to \mathbb{R}^p$ is $m \geq 1$ times continuously differentiable and non-negative, satisfying $\int_{\mathbb{R}^p} K(u) = 1$, $\int_{\mathbb{R}^p} uK(u) \, \mathrm{d}u = 0$ and (3.13), and that $f_{0,K}$ is formed by convolving f_0 with K, as in (3.12). Then if (3.6) holds:

(ii) *$f_{0,K}$ is invertible, bi-Lipschitz, and m times continuously differentiable.*

4 Application: a nonlinear Phillips curve?

4.1 Formulation as an endogenous regime-switching SVAR

The inflation surge that followed the COVID-19 pandemic reignited academic interest in possible nonlinearity in the transmission of supply shocks to inflation. However, views on the relevance of nonlinearity are divided. On the one hand, Ball et al. (2022) and Benigno and Eggertsson (2023) find significant nonlinearity in their formulations of the Phillips curve, and argue that these nonlinearities are needed to account for the recent inflation surge. On the other hand, Beaudry et al. (2025) caution that the evidence on nonlinearity is not robust to functional form assumptions, especially as pertains to the treatment of expectations. Reconsidering this debate, through the lens of a endogenous regime-switching SVAR, provides an illustrative application of the methodology developed in this paper.

Our identification result can be useful in this debate because it highlights the following: since all observationally equivalent (possibly nonlinear) structures are identified up to a (linear) orthogonal transformation, then if one finds no (statistically significant) evidence of nonlinearity under one specific identification scheme, then this will remain true however the model is identified. Indeed, one can see from the orthogonal reduced-form parametrisation developed in Section 2.3 above, that the structural parameters f_0 (and $\boldsymbol{f}_1$) will be nonlinear if and only its normalised (and exactly identified) counterpart g_0 (and $\boldsymbol{g}_1$) is also nonlinear, as will be the case for Qf_0 (and $Q\boldsymbol{f}_1$) for *any* $Q \in \mathbb{O}(p)$. Thus the presence of nonlinearity can be tested for in a way that is robust to the identifying scheme employed. To be clear, this is a consequence of modelling the joint determination of $z_t = (\log \theta_t, \pi_t)^\intercal$ in its entirety; the argument does not carry over to the methodology employed in the aforementioned papers, because these provide only a single-equation analysis of the Phillips curve, and so their findings are potentially contingent on the assumptions made in order identify that equation.

Building on the development already given to this point in Example 2.1, inspired by the recent work of Benigno and Eggertsson (2023) we consider the following endogenous regime switching SVAR for $z_t = (\log \theta_t, \pi_t)^\intercal$,

$$\Phi_0^{(\ell_t)} z_t = c + \sum_{j=1}^{p} \Phi_j^{(\ell_{t-j})} z_{t-j} + \varepsilon_t, \qquad \varepsilon_t \sim_{\text{i.i.d.}} N[0, I_k] \tag{4.1}$$

where $\theta_t = v_t/u_t$ is the vacancy–unemployment ratio, π_t is consumer price inflation, ε_t are the structural shocks, and

$$\ell_t := \begin{cases} 1 & \text{if } z_{1t} = \log \theta_t \leq 0 \quad \text{(normal labour market)} \\ 2 & \text{if } z_{1t} = \log \theta_t > 0 \quad \text{(labour shortage).} \end{cases} \tag{4.2}$$

This model thus has two regimes, determined by the sign of z_{1t}. Following the arguments that lead to (3.7) above, to ensure continuity of the model in both z_t and its lags, we parametrise the regime-dependent coefficient matrices non-redundantly as

$$\Phi_i^{(\ell)} = \begin{bmatrix} \Phi_{i,11}^{(\ell)} & \Phi_{i,12}^{(\ell)} \\ \Phi_{i,21}^{(\ell)} & \Phi_{i,22}^{(\ell)} \end{bmatrix} = \begin{bmatrix} \Phi_{i,11}^{(\ell)} & \Phi_{i,12} \\ \Phi_{i,21}^{(\ell)} & \Phi_{i,22} \end{bmatrix}, \ \ell \in \{1, 2\} \tag{4.3}$$

so that only the coefficients of the regime-determining variable, z_{1t}, are permitted to vary across the two regimes. The model is then guaranteed to yield a solution for z_t, for every possible value of the r.h.s., provided that $\det \Phi_0^{(1)} \det \Phi_0^{(2)} > 0$.

To obtain a just-identified specification, by Theorem 2.2 it suffices to impose $p(p-1)/2 = 1$ restrictions on the model parameters (see also the discussion in Section 2.3 above). For some identifying schemes, this may involve imposing a restriction on only one of the two regimes. However, the identifying assumption in Benigno and Eggertsson (2023) corresponds to the 'recursive' or 'Cholesky' restriction under which inflation $\pi_t = z_{2t}$ has no contemporaneous effect on tightness $\log \theta_t = z_{1t}$, and thus that the matrix $\Phi_0^{(\ell)}$ is lower triangular for $\ell \in \{1, 2\}$. In view of (4.3), this in fact constitutes only a single restriction on the model parameters, that $\Phi_{0,12} = 0$, and so is exactly rather than over-identifying. The second equation of the nonlinear SVAR (4.1) can in this case be estimated by nonlinear regression (with π_t as the dependent variable), as was done by Benigno and Eggertsson (2023).

4.2 Testing for linearity in the Phillips curve

Let $\{\tilde{\Phi}_i^{(\ell)}\}$ denote the coefficients corresponding to the recursive identification scheme of Benigno and Eggertsson (2023). In light of Section 2.3, these are the coefficients associated with the orthogonal reduced form representation (2.7) when g_j are modelled as piecewise linear. Theorem 2.2, together with a sign-normalisation of the shocks, then implies that *all* observationally equivalent models can be obtained by a *common* rotation of the recursively identified model, i.e. $\Phi_i^{(\ell)} = Q\tilde{\Phi}_i^{(\ell)}$ for $\ell \in \{1, 2\}$ and $i \in \{0, \dots k\}$, where $Q \in \mathbb{O}(p)$ with $\det Q > 0$. Hence some regime dependence in $\{\tilde{\Phi}_i^{(\ell)}\}$ is a necessary condition for the existence of a nonlinear Phillips curve under *any* identification scheme. Since the null hypothesis of no regime dependence in $\{\tilde{\Phi}_i^{(\ell)}\}$ is testable, a failure to reject it would provide evidence, in favour of a linear Phillips curve, that is robust to all possible identifying schemes. (In this respect, our imposition of the Benigno and Eggertsson, 2023, restrictions merely provides a convenient way to normalise the system, in the manner of Section 2.3).

Observe that the specification of (4.1) allows for nonlinearities in all lags of the SVAR. This permits the dynamic response of inflation to labour market tightness shocks to be nonlinear, even if the impact responses are linear, i.e., even if $\Phi_0^{(\ell)}$ is regime-invariant. We therefore consider two separate tests of linearity. The first tests

$$H_0^{\mathrm{NS}} : \Phi_0^{(1)} = \Phi_0^{(2)} \quad \text{against} \quad H_1^{\mathrm{NS}} : \Phi_0^{(1)} \neq \Phi_0^{(2)}. \tag{4.4}$$

The null hypothesis H_0^{NS} can be interpreted as saying that there is no *endogenous* regime switching, and implies that the impact effect of labour tightness shocks on inflation does not depend on the state of the labour market. However, H_0^{NS} does not exclude the possibility that $\Phi_i^{(1)} \neq \Phi_i^{(2)}$ for some $i \in \{1, \dots, k\}$, in which case the dynamic effects of tightness shocks may still be regime dependent, at longer horizons.

Our second hypothesis is more restrictive:

$$H_0^{\mathrm{lin}} : \Phi_i^{(1)} = \Phi_i^{(2)} \quad \forall i \in \{0, \dots, k\} \quad \text{against} \quad H_1^{\mathrm{lin}} : \Phi_i^{(1)} \neq \Phi_i^{(2)} \quad \text{for some } i. \tag{4.5}$$

Table 4.1: Likelihood Ratio Tests of linearity hypotheses

Null Hypothesis	Restrictions	LR Statistic [p-value]	
		1960Q1–2024Q4	2008Q1–2024Q4
No Endogenous Switching (4.4)	2	21.7 [0.0]	38.6 [0.0]
Linear SVAR (4.5)	6	38.0 [0.0]	51.2 [0.0]

Notes: The model is a bivariate SVAR in log vacancy–unemployment ratio ($\log \theta$) and wage inflation with two lags and two regimes, determined by the sign of ($\log \theta$). Both tests are against the alternative of an unrestricted piecewise linear SVAR. Asymptotic p-values based on χ^2 distribution with degrees of freedom equal to the number of restrictions.

Here the null hypothesis H_0^{lin} entails a linear SVAR. Failure to reject H_0^{lin} would suggest that a linear SVAR provides an adequate description of the dynamic causal effects (modulo the usual invertibility caveats), and thus that the Phillips curve is linear under any identification scheme.

4.3 Results

We use the data from the 2025 version of Benigno and Eggertsson (2023), available on the authors' websites. Specifically, inflation π_t is the quarterly, annualised core CPI inflation (excluding food and energy), constructed from monthly CPI data averaged to quarterly frequency and sourced from the BLS via FRED. The vacancy-to-unemployment ratio $\theta_t = v_t/u_t$ is the ratio of job vacancies to unemployed workers, using the Barnichon (2010) vacancy series (as updated by the author), also averaged from a monthly to a quarterly frequency. We estimate the piecewise linear SVAR (4.1) with two lags over the sample periods 1960Q1-2024Q4 and 2008Q1-2024Q4, to mirror the analysis of Benigno and Eggertsson (2023).

4.3.1 Testing linearity

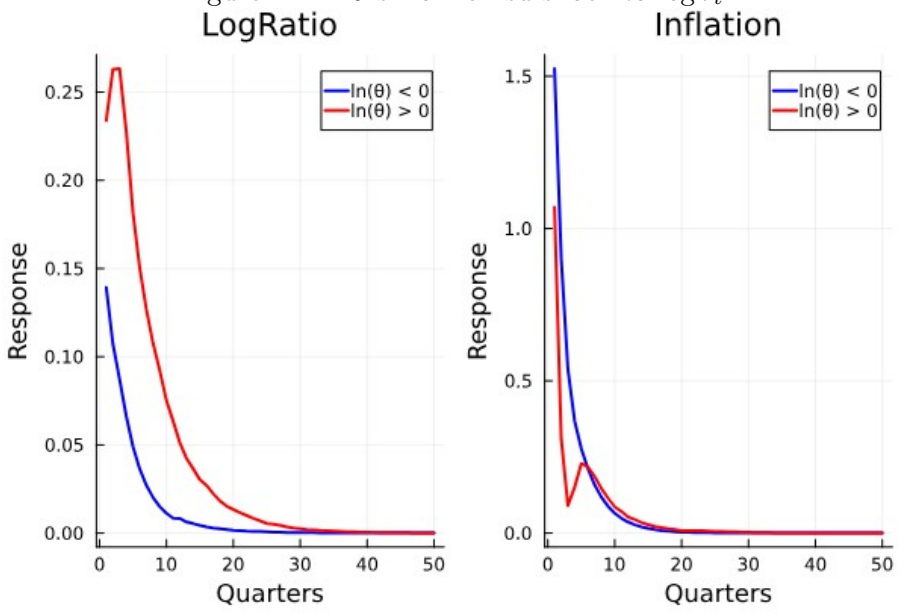
Table 4.1 reports likelihood ratio (LR) tests of our two linearity hypotheses: H_0^{NS} (no endogenous regime switching) and H_0^{lin} (fully linear SVAR). The results clearly reject the linearity hypothesis, both in its weak and strong forms. The apparent deterioration in fit of the linear models is even stronger in the shorter, more recent, sample.

Even though failure to reject would have been conclusive evidence against nonlinearity, these results are not enough to conclude that the Phillips curve itself is nonlinear. They imply that impulse responses to identified structural inflation and tightness shocks will be significantly state-dependent under any identification scheme, but it remains to be seen what this state dependence looks like under any specific identification scheme. We turn to this question next.

4.3.2 Impulse responses

To highlight different policy trade-offs across regimes, we compute state-dependent impulse response functions following a positive 1 standard deviation shock to $\log \theta_t$. We choose two starting points in the sample, which differ by the sign of $\log \theta_t$. Figure 4.1 shows that a positive 1 standard deviation shock to labor market tightness increases $\log \theta_t$ and inflation π_t under both regimes. However, the effect under $\log \theta_t > 0$ is larger on market tightness and smaller on inflation, which implies a reduced inflation-output trade-off compared to the $\log \theta_t < 0$ regime.

Figure 4.1: IRFs from a 1sd shock to $\log \theta_t$.



Note: The blue series shows IRFs from a starting point characterized by $\log \theta_t < 0$, while the red series shows IRFs from a starting point characterized by $\log \theta_t > 0$.

4.3.3 Phillips curve slope

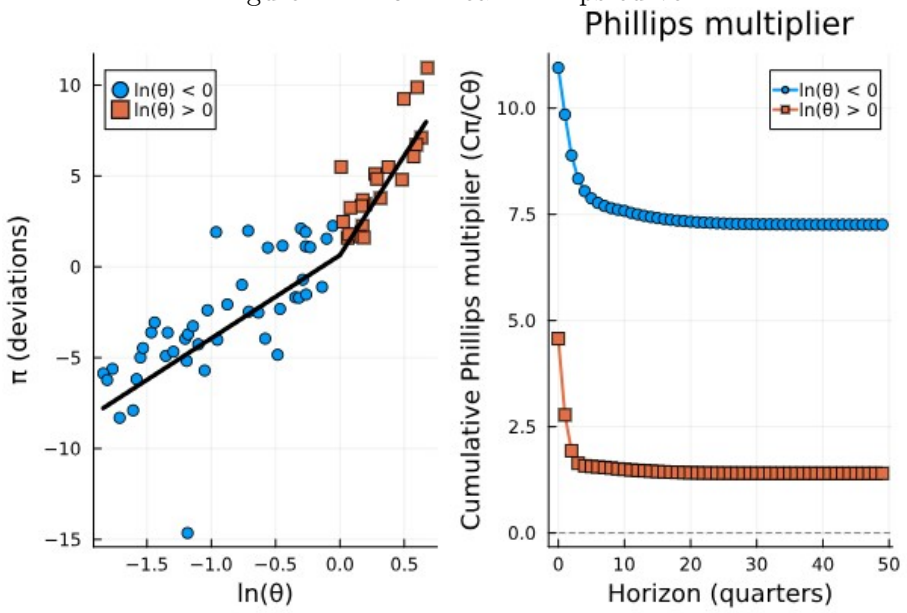
Further evidence on the non-linear Phillips curve is obtained by computing estimates of its slope under both regimes. We do this in two different ways. First, we produce a kinked Phillips curve plot (equivalent to Figure 6(b) of Benigno and Eggertsson, 2023). This is shown in the left panel of Figure 4.2. The scatterplot shows inflation after removing the effects of all explanatory variables from the supply equation in model (4.1) except $\log \theta_t$. The solid lines trace out the estimated Phillips curve in the $(\log \theta_t, \pi_t)$ space. In particular, the slope coefficient under each regime is computed as $-\Phi_{0,21}^{(\ell)}/\Phi_{0,22}$, which is given by the equation in the bottom row of (4.1), solved for $z_{2,t} = \pi_t$, and using the fact that $\Phi_{0,22}$ is regime-independent, see (4.3).

The right panel of Figure 4.2 shows a dynamic Phillips curve multiplier under each regime by taking ratios of the cumulative inflation IRF (at horizon h) following a market tightness shock which raises $\log \theta_t$ by 1 unit over the next h periods:

$$\text{Slope}_h^{PC} = \frac{\sum_{s=0}^{h} \frac{\partial \pi_{t+s}}{\partial \varepsilon_{\theta,t}}}{\sum_{s=0}^{h} \frac{\partial \log(\theta_{t+s})}{\partial \varepsilon_{\theta,t}}}.$$

Both approaches show a substantially steeper Phillips curve in the $\log \theta_t > 0$ regime compared to the $\log \theta_t \leq 0$ regime. The results are qualitatively and quantitatively consistent with Benigno and Eggertsson (2023), which is not surprising given that we used the same identifying assumption as them.

Figure 4.2: Non-linear Phillips curve.



Left panel: scatter plot of 'inflation deviations' and $\log(\theta)$, sample 2008 Q1- 2024 Q4. *Right panel:* Dynamic Phillips curve multiplier by taking ratios of cumulative IRFs, sample 2008 Q1- 2024 Q4.

References

Arias, J. E., J. F. Rubio-Ramirez, and D. F. Waggoner (2018): "Inference based on structural vector autoregressions identified with sign and zero restrictions: theory and applications," *Econometrica*, 86, 685–720.

Aruoba, S. B., M. Mlikota, F. Schorfheide, and S. Villalvazo (2022): "SVARs with occasionally-binding constraints," *Journal of Econometrics*, 231, 477–499.

Auerbach, A. J. and Y. Gorodnichenko (2012): "Measuring the output responses to fiscal policy," *American Economic Journal: Economic Policy*, 4, 1–27.

Ball, L., D. Leigh, and P. Mishra (2022): "Understanding US inflation during the COVID-19 era," *Brookings Papers on Economic Activity*, 2022, 1–80.

Barnichon, R. (2010): "Building a composite help-wanted index," *Economics Letters*, 109, 175–178.

Beaudry, P., C. Hou, and F. Portier (2025): "On the fragility of the nonlinear Phillips curve view of recent inflation," National Bureau of Economic Research, Working Paper 33522.

Benigno, P. and G. B. Eggertsson (2023): "It's baaack: The surge in inflation in the 2020s and the return of the non-linear Phillips curve," National Bureau of Economic Research, Working Paper 31197.

Berry, S. T. and P. A. Haile (2018): "Identification of nonparametric simultaneous equations models with a residual index structure," *Econometrica*, 86, 289–315.

Bingham, N. H. (2001): "Random walk and fluctuation theory," *Handbook of Statistics*, 19, 171–213.

Bruns, M. and M. Piffer (2024): "Tractable Bayesian estimation of smooth transition vector autoregressive models," *Econometrics Journal*, 27, 343–361.

Caggiano, G., E. Castelnuovo, V. Colombo, and G. Nodari (2015): "Estimating fiscal multipliers: News from a non-linear world," *Economic Journal*, 125, 746–776.

Carriero, A., T. E. Clark, M. Marcellino, and E. Mertens (2025): "Forecasting with shadow rate VARs," *Quantitative Economics*, 16, 795–822.

Chan, K. S., ed. (2009): *Exploration of a Nonlinear World: an appreciation of Howell Tong's contributions to statistics*, World Scientific.

Chernozhukov, V., A. Galichon, M. Henry, and B. Pass (2021): "Identification of hedonic equilibrium and nonseparable simultaneous equations," *Journal of Political Economy*, 129, 842–870.

Deimling, K. (1985): *Nonlinear Functional Analysis*, Springer.

Duffy, J. A. and S. Mavroeidis (2024): "Common trends and long-run identification in nonlinear structural VARs," arXiv:2404.05349.

Duffy, J. A., S. Mavroeidis, and S. Wycherley (2023): "Stationarity with Occasionally Binding Constraints," arXiv:2307.06190.

Evans, L. C. and R. F. Gariepy (2015): *Measure Theory and Fine Properties of Functions*, CRC Press, revised ed.

Faust, J. (1998): "The robustness of identified VAR conclusions about money," *Carnegie-Rochester Conference Series on Public Policy*, 49, 207–244.

Friesecke, G., R. D. James, and S. Müller (2002): "A theorem on geometric rigidity and the derivation of nonlinear plate theory from three-dimensional elasticity," *Communications on Pure and Applied Mathematics*, 55, 1461–1506.

Gao, J. and P. C. B. Phillips (2013): "Semiparametric estimation in triangular system equations with nonstationarity," *Journal of Econometrics*, 176, 59–79.

Gouriéroux, C., J. J. Laffont, and A. Monfort (1980): "Coherency conditions in simultaneous linear equation models with endogenous switching regimes," *Econometrica*, 48, 675–695.

Gouriéroux, C., A. Monfort, and J.-P. Renne (2020): "Identification and estimation in non-fundamental structural VARMA models," *Review of Economic Studies*, 87, 1915–1953.

Hamilton, J. D. (1994): *Time Series Analysis*, Princeton University Press.

HUBRICH, K. AND T. TERÄSVIRTA (2013): "Thresholds and smooth transitions in vector autoregressive models," in *VAR Models in Macroeconomics – New Developments and Applications: essays in honor of Christopher A. Sims*.

JOHN, F. (1961): "Rotation and strain," *Communications on Pure and Applied Mathematics*, 14, 391–413.

KILIAN, L. AND H. LÜTKEPOHL (2017): *Structural Vector Autorgressive Analysis*, C.U.P.

LANNE, M., M. MEITZ, AND P. SAIKKONEN (2017): "Identification and estimation of non-Gaussian structural vector autoregressions," *Journal of Econometrics*, 196, 288–304.

LÜTKEPOHL, H. (2007): *New Introduction to Multiple Time Series Analysis*, Springer, 2nd ed.

MATZKIN, R. L. (2008): "Identification in nonparametric simultaneous equations models," *Econometrica*, 76, 945–978.

——— (2015): "Estimation of nonparametric models with simultaneity," *Econometrica*, 83, 1–66.

MAVROEIDIS, S. (2021): "Identification at the zero lower bound," *Econometrica*, 69, 2855–2885.

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

RUBIO-RAMIREZ, J. F., D. F. WAGGONER, AND T. ZHA (2005): "Markov-switching structural vector autoregressions: theory and application," Working Paper 2005-27.

——— (2010): "Structural vector autoregressions: theory of identification and algorithms for inference," *Review of Economic Studies*, 77, 665–696.

SCHOLTES, S. (2012): *Introduction to Piecewise Differentiable Equations*, Springer.

SIMS, C. A. (1980): "Macroeconomics and reality," *Econometrica*, 1–48.

SIMS, C. A. AND T. ZHA (2006): "Were there regime switches in US monetary policy?" *American Economic Review*, 96, 54–81.

STOCK, J. H. AND M. W. WATSON (2018): "Identification and estimation of dynamic causal effects in macroeconomics using external instruments," *Economic Journal*, 128, 917–948.

TERÄSVIRTA, T., D. TJØSTHEIM, AND C. W. J. GRANGER (2010): *Modelling Nonlinear Economic Time Series*, O.U.P.

UHLIG, H. (1998): "The robustness of identified VAR conclusions about money: a comment," *Carnegie-Rochester Conference Series on Public Policy*, 49, 245–263.

——— (2005): "What are the effects of monetary policy on output? Results from an agnostic identification procedure," *Journal of Monetary Economics*, 52, 381–419.

A Proofs of identification results

A.1 Reformulation of the problem

While the nonlinear SVAR of Section 2 is a (dynamic) simultaneous equations model (SEM), our notion of observational equivalence refers only the distribution of z_t conditional on its lags. This allows the proof of Theorem 2.2 to be approached in a manner that entirely abstracts from the dynamics of the SVAR. To more clearly connect our underlying identification results with those of the literature on nonlinear simultaneous equations models (SEMs), in particular Matzkin (2008), in this appendix we consider the nonlinear SEM

$$U = r(Y, X) = r_0(Y) + r_1(X), \tag{A.1}$$

where U and Y are random vectors taking values in $\mathbb{R}^G$, and X is a random vector taking values in $\mathbb{R}^K$. Let f_U denote the density of U, location- and scale-normalised so that $\mathbb{E}U = 0$ and $\mathbb{E}UU^\intercal = I_G$. This is the same model as in (2.1) of Matzkin (2008), but with the additional restriction that r is (additively) separable in the endogenous and exogenous variables, Y and X. Our results on observational equivalence in this model are given as Theorem A.1 below: on the basis of which the proof of Theorem 2.2 will simply be a matter of translating between the notation of the SVAR in Section 2, and that of (A.1) (see Appendix A.4 below).

Under the regularity conditions imposed below on the functions $\Gamma_0 \ni r_0$, if we suppose that X has Lebesgue density f_X with support $\mathbb{R}^K$, then the model implies that the distribution of Y conditional on X has a Lebesgue density that satisfies (see e.g. Evans and Gariepy, 2015, Thm. 3.9)

$$f_{Y|X}(y \mid x) = f_U[r(y,x)] \cdot \det Dr_0(y) = f_U[r_0(y) + r_1(x)] \cdot \det Dr_0(y)$$

a.e. $(y, x) \in \mathbb{R}^{G+K}$; here the 'a.e.' qualifier is a consequence both of the usual non-uniqueness of the conditional density (with respect to modifications on a null set), and more importantly the fact the Jacobian $Dr_0(y)$ need only exist a.e. We will accordingly say that two alternative parametrisations $(\tilde{r}_0, \tilde{r}_1, f_{\tilde{U}})$ and (r_0, r_1, f_U) are *observationally equivalent* if

$$f_U[r(y,x)] \cdot \det Dr_0(y) = f_U[\tilde{r}(y,x)] \cdot \det D\tilde{r}_0(y) \tag{A.2}$$

a.e. $(y, x) \in \mathbb{R}^{G+K}$, i.e. if they imply the same density for Y conditional on X. (This accords exactly with the definition of observational equivalence given in Section 2.2, transposed from the nonlinear SVAR to the nonlinear SEM.)

The model is parametrised by the functions $r_0 : \mathbb{R}^G \to \mathbb{R}^G$, $r_1 : \mathbb{R}^K \to \mathbb{R}^G$, and the density f_U. Let $\Gamma_i \ni r_i$, for $i \in \{0, 1\}$, and $\Phi \ni f_U$ denote the sets of functions and densities that together comprise the model parameter space. We make only weak assumptions on the elements of those parameter spaces, and some further assumptions on the parameters (r_0, r_1, f_U) that actually generated the data; for a discussion of these conditions, see Section 2.2.

Assumption SEM. *Γ_0, Γ_1 and Φ collect every function such that:*

1. *$\tilde{r}_0 \in \Gamma_0$ and $\tilde{r}_1 \in \Gamma_1$ are locally Lipschitz (continuous).*

2. $\tilde{r}_0 \in \Gamma_0$ *is a bijection* $\mathbb{R}^G \to \mathbb{R}^G$*, with* $\det D\tilde{r}_0(y) > 0$ *for almost every* $y \in \mathbb{R}^G$*.*

3. $f_{\tilde{U}} \in \Phi$ *is continuously differentiable, with* $f_{\tilde{U}}(u) > 0$ *for all* $u \in \mathbb{R}^G$*, and*

$$\int_{\mathbb{R}^G} f_{\tilde{U}}(u)\,\mathrm{d}u = 1 \qquad \int_{\mathbb{R}^G} u f_{\tilde{U}}(u)\,\mathrm{d}u = 0, \qquad \int_{\mathbb{R}^G} uu^{\mathsf{T}} f_{\tilde{U}}(u)\,\mathrm{d}u = I_G.$$

(r_0, f_1, f_U) *are such that:*

4. $r_1 : \mathbb{R}^K \to \mathbb{R}^G$ *is surjective, with* $\operatorname{rk} Dr_1(x) = G$ *for almost every* $x \in \mathbb{R}^K$*;*

5. r_0^{-1} *is locally Lipschitz; and*

6. f_U *has a* local *maximum at some* $u^* \in \mathbb{R}^G$*, and is twice continuously differentiable in a neighbourhood of* u^**, with negative definite Hessian there.*

We can now state our main result on observational equivalence in the model (A.1). Recall that $\mathbb{O}(p)$ denote the set of $p \times p$ orthogonal matrices; further define $\mathbb{O}^+(p)$ to be the subset of these matrices with positive determinant.

Theorem A.1. *Suppose that* SEM *holds. Let* $\tilde{r}_i \in \Gamma_i$ *for* $i \in \{0, 1\}$*. Then there exists an* $f_{\tilde{U}} \in \Phi$ *such that* $(\tilde{r}_0, \tilde{r}_1, f_{\tilde{U}})$ *is observationally equivalent to* (r_0, r_1, f_U)*, if and only if there exists a* $Q \in \mathbb{O}^+(G)$ *such that*

$$\tilde{r}_0(y) + \tilde{r}_1(x) = Q[r_0(y) + r_1(x)] \tag{A.3}$$

for all $(y, x) \in \mathbb{R}^G \times \mathbb{R}^K$*.*

Only the sum of $r_0(y) + r_1(x)$ is identified, because in view of (A.1) we cannot distinguish between (r_0, r_1) and $(r_0 - \delta, r_1 + \delta)$ for any $\delta \in \mathbb{R}^G$. This indeterminacy can of course be resolved by imposing a location normalisation on either of these functions, e.g. by requiring $\tilde{r}_0(0) = 0$ for all $\tilde{r}_0 \in \Gamma_0$.

A.2 Preliminaries

For ease of reference, the following lemma collects some useful (and well known) results regarding the properties of locally Lipschitz functions, that will be relied on in the proof. Note when we say that a function $g : \mathbb{R}^k \to \mathbb{R}^\ell$ is differentiable at $x_0 \in \mathbb{R}^k$, we mean that there exists a (Jacobian) matrix $Dg(x_0) \in \mathbb{R}^{\ell \times k}$, such that

$$g(x) - g(x_0) = Dg(x_0)(x - x_0) + o(\|x - x_0\|)$$

as $x \to x_0$. When we refer to the 'measure' of a subset of Euclidean space, we always mean its Lebesgue measure.

Lemma A.1. *Suppose that* $g : \mathbb{R}^k \to \mathbb{R}^\ell$ *is locally Lipschitz. Then*

(i) g *is differentiable a.e.;*

(ii) *if* $k \leq \ell$*, and* $N \subseteq \mathbb{R}^k$ *has measure zero (in* $\mathbb{R}^k$*), then* $g(N)$ *has measure zero (in* $\mathbb{R}^\ell$*);*

(iii) *if* $Dg(x) = B$ *for almost every* $x \in \mathbb{R}^k$*, then* $g(x) = a + Bx$ *for all* $x \in \mathbb{R}^k$*; and*

(iv) *if* $\ell = k \geq 2$*, and* $Dg(x) \in \mathbb{O}^+(k)$ *for almost every* $x \in \mathbb{R}^k$*, then* $g(x) = a + Qx$ *for some* $Q \in \mathbb{O}^+(k)$*.*

Suppose that $k = \ell$*,* g *is bijective, and* g^{-1} *and* $h : \mathbb{R}^k \to \mathbb{R}^m$ *are locally Lipschitz. Then*

(v) *for almost every* $x \in \mathbb{R}^k$*,* $f \coloneqq h \circ g$ *is differentiable at* x*, and*

$$Df(x) = Dh[g(x)]Dg(x).$$

Proof. (i). This is Rademacher's theorem (e.g. Theorem 3.2 in Evans and Gariepy, 2015).

(ii). This follows by Lemma 2.2(i), Theorem 2.5 and and Theorem 2.8(i) in Evans and Gariepy (2015).

(iii). Fix $x_0 \in \mathbb{R}^k$. Since the locally Lipschitz function $f(x) \coloneqq g(x) - Bx$ has $Df(x) = 0$ a.e., and is absolutely continuous along the segment joining any point $x \in \mathbb{R}^k$ to x_0, it must be constant along that segment, by the fundamental theorem of calculus. Hence $f(x) = f(x_0) =: a$ for all x.

(iv). This follows from Theorem 3.1 (and the discussion on p. 1469) in Friesecke, James, and Müller (2002) – see also Theorem IV in John (1961) – and part (iii).

(v). Let $G \subseteq \mathbb{R}^k$ and $H \subseteq \mathbb{R}^k$ collect the points at which g and h are respectively differentiable. Then $\mathbb{R}^k \backslash H$ has measure zero, and since g^{-1} is surjective and locally Lipschitz, it follows from $\mathbb{R}^k = g^{-1}(\mathbb{R}^k \backslash H) \cup g^{-1}(H)$ and part (ii) that $\mathbb{R}^k \backslash g^{-1}(H)$ also has measure zero. Deduce that the complement of $X \coloneqq G \cap g^{-1}(H)$ has measure zero, and that for every $x \in X$, g is differentiable at x, and h is differentiable at $g(x)$. Thus the chain rule yields the result. □

A.3 Proof of Theorem A.1

It is clear that if (A.3) holds, then

$$\tilde{U} \coloneqq \tilde{r}_0(Y) + \tilde{r}_1(X) = Q[r_0(Y) + r_1(X)] = QU$$

will be independent of X, with a density $f_{\tilde{U}}$ that satisfies SEM.3; hence observational equivalence obtains in this case. It remains therefore to prove the reverse implication.

To that end, we suppose that $(\tilde{r}_0, \tilde{r}_1, f_{\tilde{U}})$ is observationally equivalent to (r_0, r_1, f_U). Taking logs in (A.2), as we may under SEM.2–3, yields that

$$\log f_U[r(y,x)] - \log f_{\tilde{U}}[\tilde{r}(y,x)] = \log\det D\tilde{r}_0(y) - \log\det Dr_0(y) \tag{A.4}$$

(y,x)-a.e. In view of SEM.1–2 and SEM.4, we may define a set $\mathcal{A} \subset \mathbb{R}^{G+K}$, whose complement has measure zero (in $\mathbb{R}^{G+K}$), such that for every $(y,x) \in \mathcal{A}$:

- (A.4) holds;
- r_0 and $\tilde{r}_0$ are differentiable at y, with $\det Dr_0(y) > 0$ and $\det D\tilde{r}_0(y) > 0$; and
- r_1 and $\tilde{r}_1$ are differentiable at x, with $\operatorname{rk} Dr_1(x) = G$.

By Tonelli's theorem, we may also define sets $\mathcal{Y} \subset \mathbb{R}^G$ and $\mathcal{X} \subset \mathbb{R}^K$, whose complements (in $\mathbb{R}^G$ and $\mathbb{R}^K$ respectively) have measure zero, such that:

- for every $y_0 \in \mathcal{Y}$: $(y_0, x) \in \mathcal{A}$ for almost every $x \in \mathbb{R}^K$; and
- for every $x_0 \in \mathcal{X}$: $(y, x_0) \in \mathcal{A}$ for almost every $y \in \mathbb{R}^G$.

The proof now proceeds in five steps. (Had we made imposed the stronger requirement that $\tilde{r}_0$ and $\tilde{r}_1$ be twice continuously differentiable, then the claims proved in the first two steps would follow more directly as corollaries to the results of Matzkin (2008), particularly her Theorem 3.3; and indeed our arguments in those parts of the proof largely follow hers, suitably modified to allow $\tilde{r}_0$ and $\tilde{r}_1$ to have points of non-differentiability.)

(i) Claim: $\operatorname{rk} D\tilde{r}_1(x) = G$ **for all** $x \in \mathcal{X}$**.**

Let $x_0 \in \mathcal{X}$ be given. Differentiating both sides of (A.4) with respect to x, we obtain

$$D(\log f_U)[r(y, x_0)]Dr_1(x_0) = D(\log f_{\tilde{U}})[\tilde{r}(y, x_0)]D\tilde{r}_1(x_0) \tag{A.5}$$

y-a.e. By the continuity of both sides (in y), this holds for all $y \in \mathbb{R}^G$. Recall that $\operatorname{rk} Dr_1(x_0) = G$ by the definition of $\mathcal{X}$; we must show that this is transmitted to $D\tilde{r}_1(x_0)$.

Under SEM.6, it follows from the inverse function theorem that the map

$$u \mapsto D(\log f_U)(u)^\mathsf{T}$$

is invertible in a neighbourhood of $u = u^*$, and equals zero at u^*. Hence by SEM.2, the composite map

$$y \mapsto s(y, x_0) \coloneqq D(\log f_U)[r(y, x_0)]^\mathsf{T}$$

is also invertible for y in a neighbourhood of

$$y^*(x_0) \coloneqq r_0^{-1}[u^* - r_1(x_0)], \tag{A.6}$$

with the property that

$$s[y^*(x_0), x_0] = D(\log f_U)(u^*)^\mathsf{T} = 0.$$

Hence there exist $\lambda > 0$ and $\{y^i\}_{i=1}^G$ such that

$$s(y^i, x_0) = \lambda e_i$$

for all $i \in \{1, \ldots, G\}$, where e_i denotes the ith column of I_G. Evaluating (A.5) at each y^i, we obtain that

$$Dr_1(x_0)^\mathsf{T} e_i \subset \operatorname{sp} D\tilde{r}_1(x_0)^\mathsf{T}$$

for $i \in \{1, \ldots, G\}$, whence $\operatorname{sp} Dr_1(x_0)^\mathsf{T} \subset \operatorname{sp} D\tilde{r}_1(x_0)^\mathsf{T}$. Since $Dr_1(x_0)^\mathsf{T}$ has rank G, it follows that so too does $D\tilde{r}_1(x_0)^\mathsf{T}$.

(ii) Claim: $\log\det D\tilde{r}_0(y) - \log\det Dr_0(y)$ **is constant on** $\mathcal{Y}$**.**

Let $J : \mathbb{R}^G \to \mathbb{R}^G$ be defined such that

$$J(y) = \log\det D\tilde{r}_0(y) - \log\det Dr_0(y)$$

for all $y \in \mathcal{Y}$, so that it equals the r.h.s. of (A.4) there; and set $J(y) = 0$ otherwise.

Consider again the map $y^* : \mathbb{R}^K \to \mathbb{R}^G$, defined in (A.6) above, which is surjective and locally Lipschitz in view of SEM.1 and SEM.4–5. Hence the complement of $y^*(\mathcal{X})$ in $\mathbb{R}^G$ has measure zero, by Lemma A.1(ii). Now fix $y_0 \in \mathcal{Y} \cap y^*(\mathcal{X})$, whose complement also has measure zero. By definition of $\mathcal{Y}$, (A.4) holds at (y_0, x), for almost every x. Moreover, since both sides of (A.4) are continuous in x, it follows that

$$\log f_U[r(y_0, x)] - \log f_{\tilde{U}}[\tilde{r}(y_0, x)] = J(y_0) \tag{A.7}$$

holds for *every* $x \in \mathbb{R}^K$. Since $y_0 \in \mathcal{Y}$, the l.h.s. is differentiable with respect to y, whence so too is the r.h.s., with

$$D(\log f_U)[r(y_0, x)]Dr_0(y_0) - D(\log f_{\tilde{U}})[\tilde{r}(y_0, x)]D\tilde{r}_0(y_0) = DJ(y_0) \tag{A.8}$$

Since $y_0 \in y^*(\mathcal{X})$, there exists an $x_0 \in \mathcal{X}$ such that $r_0(y_0, x_0) = u^*$, and hence

$$D(\log f_U)[r(y_0, x_0)] = D(\log f_U)(u^*) = 0.$$

Since (A.5) holds at (y_0, x_0), with $\operatorname{rk} D\tilde{r}_1(x_0) = \operatorname{rk} Dr_1(x_0) = G$ by the preceding part of the proof, it follows that

$$D(\log f_{\tilde{U}})[\tilde{r}(y_0, x_0)] = 0.$$

Deduce from (A.8) that $DJ(y_0) = 0$ for all $y_0 \in \mathcal{Y} \cap y^*(\mathcal{X})$. It then follows from (A.7) above that for all $x \in \mathbb{R}^K$, the Jacobian of

$$y \mapsto \log f_U[r(y, x)] - \log f_{\tilde{U}}[\tilde{r}(y, x)]$$

is zero at $y_0 \in \mathcal{Y} \cap y^*(\mathcal{X})$, i.e. almost everwhere. Since this map is locally Lipschitz, it is therefore equal to some constant C, by Lemma A.1(iii). Hence

$$J(y) = \log\det D\tilde{r}_0(y) - \log\det Dr_0(y) = C$$

for all $y \in \mathcal{Y}$.

(iii) Claim: $\tilde{r}_1(x) = \tilde{u}^* - \tilde{m}_0[u^* - r_1(x)]$**, for** $\tilde{m}_0 \coloneqq \tilde{r}_0 \circ r_0^{-1}$**.**

Returning now to (A.4) above, it follows from the preceding part of the proof that

$$\log f_{\tilde{U}}[\tilde{r}_0(y) + \tilde{r}_1(x)] = \log f_U[r_0(y) + r_1(x)] + C$$

(y,x)-a.e.; and since both sides are continuous in (y,x), the preceding must hold for *all* $(y,x) \in \mathbb{R}^{G+K}$. Setting $\tilde{u} = \tilde{r}_0(y) + \tilde{r}_1(x)$, and recalling that $\tilde{r}_0$ is invertible (by SEM.2), this may be equivalently stated as

$$\begin{aligned}\log f_{\tilde{U}}(\tilde{u}) &= \log f_U[r_0\{\tilde{r}_0^{-1}[\tilde{u} - \tilde{r}_1(x)]\} + r_1(x)] + C \\ &= \log f_U\{\tilde{m}_0^{-1}[\tilde{u} - \tilde{r}_1(x)] + r_1(x)\} + C\end{aligned}$$

for all $(\tilde{u},x) \in \mathbb{R}^{G+K}$, where $\tilde{m}_0 \coloneqq \tilde{r}_0 \circ r_0^{-1}$ is invertible and locally Lipschitz, by SEM.1 and SEM.5. Since the l.h.s. of the preceding does not depend on x, the r.h.s. must be invariant to x, and so we have in particular that

$$f_U\{\tilde{m}_0^{-1}[\tilde{u} - \tilde{r}_1(0)] + r_1(0)\} = f_U\{\tilde{m}_0^{-1}[\tilde{u} - \tilde{r}_1(x)] + r_1(x)\} \tag{A.9}$$

for all $(\tilde{u},x) \in \mathbb{R}^{G+K}$.

By taking $\tilde{u}$ in the preceding to be equal to

$$\tilde{u}^* \coloneqq \tilde{m}_0[u^* - r_1(0)] + \tilde{r}_1(0), \tag{A.10}$$

for u^* as in SEM.6, we obtain that

$$f_U(u^*) = f_U\{\tilde{m}_0^{-1}[\tilde{u}^* - \tilde{r}_1(0)] + r_1(0)\} = f_U\{\tilde{m}_0^{-1}[\tilde{u}^* - \tilde{r}_1(x)] + r_1(x)\} \tag{A.11}$$

for all $x \in \mathbb{R}^K$. Defining the continuous map $\theta : \mathbb{R}^K \to \mathbb{R}^G$,

$$\theta(x) \coloneqq \tilde{m}_0^{-1}[\tilde{u}^* - \tilde{r}_1(x)] + r_1(x)$$

which by (A.10) has $\theta(0) = u^*$, we may thus rewrite (A.11) as

$$f_U(u^*) = f_U[\theta(0)] = f_U[\theta(x)] \tag{A.12}$$

for all $x \in \mathbb{R}^K$.

The preceding entails that $f_U[\theta(x)]$ does not in fact depend on x; we need to show that this implies that $\theta(x)$ itself is invariant to x. By a second-order Taylor expansion of f_U around $u = u^*$, in view of SEM.6, there exist $\epsilon, \eta > 0$ such that

$$|f_U(u) - f_U(u^*)| \geq \eta \|u - u^*\|^2$$

for all $\|u - u^*\| < \epsilon$. Since $x \mapsto \theta(x)$ is continuous with $\theta(0) = u^*$, the equalities in (A.12) can hold for all $x \in \mathbb{R}^K$ only if

$$\tilde{m}_0^{-1}[\tilde{u}^* - \tilde{r}_1(x)] + r_1(x) = \theta(x) = \theta(0) = u^*$$

for all $x \in \mathbb{R}^K$. Thus

$$\tilde{r}_1(x) = \tilde{u}^* - \tilde{m}_0[u^* - r_1(x)] \tag{A.13}$$

for all $x \in \mathbb{R}^K$.

(iv) Claim: $\tilde{m}_0$ is affine.

For $v \in \mathbb{R}^G$, define

$$\delta(v) \coloneqq f_U\{\tilde{m}_0^{-1}[(v + \tilde{u}^*) - \tilde{r}_1(0)] + r_1(0)\}$$

which in view of (A.11) satisfies

$$\delta(0) = f_U\{\tilde{m}_0^{-1}[\tilde{u}^* - \tilde{r}_1(0)] + r_1(0)\} = f_U(u^*). \tag{A.14}$$

Noting that (A.9) above holds for all $(\tilde{u}, x) \in \mathbb{R}^{G+K}$, it follows that by taking $\tilde{u} = v + \tilde{u}^*$ there, we obtain

$$\delta(v) = f_U\{\tilde{m}_0^{-1}[(v + \tilde{u}^*) - \tilde{r}_1(0)] + r_1(0)\} = f_U\{\tilde{m}_0^{-1}[(v + \tilde{u}^*) - \tilde{r}_1(x)] + r_1(x)\}$$

for all $x \in \mathbb{R}^K$. By the preceding part of the proof (namely, (A.13)),

$$\tilde{m}_0^{-1}[(v + \tilde{u}^*) - \tilde{r}_1(x)] = \tilde{m}_0^{-1}\{v + \tilde{m}_0[u^* - r_1(x)]\},$$

and hence

$$\delta(v) = f_U[\tilde{m}_0^{-1}\{v + \tilde{m}_0[u^* - r_1(x)]\} + r_1(x)],$$

with the r.h.s. being invariant to $x \in \mathbb{R}^K$. Since by SEM.4 the image of r_1 is the whole of $\mathbb{R}^G$, we may conclude that

$$\delta(v) = f_U\{\tilde{m}_0^{-1}[v + \tilde{m}_0(u^* - w)] + w\}$$

depends only on v, for all $w \in \mathbb{R}^G$; equivalently,

$$\delta(v) = f_U\{\tilde{m}_0^{-1}[v + \tilde{m}_0(w)] + u^* - w\} \tag{A.15}$$

for all $w \in \mathbb{R}^G$.

To establish that $\tilde{m}_0$ is affine, we shall now consider the behaviour of $\delta(v)$ in a neighbourhood of $v = 0$. We first note that $\delta(0) = f_U(u^*)$ by (A.14) above, and that by SEM.6 $f_U(u)$ admits the following second-order Taylor expansion,

$$f_U(u) - f_U(u^*) = -\tfrac{1}{2}(u - u^*)^{\mathsf{T}} H(u - u^*) + o(\|u - u^*\|^2) \tag{A.16}$$

as $u \to u^*$, where H is positive definite. We note that for $w \in \mathbb{R}^G$, $\tilde{m}_0^{-1}$ is differentiable at the value of $\tilde{m}_0(w)$ if $\tilde{m}_0 = \tilde{r}_0 \circ r_0^{-1}$ is itself differentiable at w with $\det D\tilde{m}_0(w) \neq 0$. Since $\tilde{r}_0$ and r_0^{-1} are locally Lipschitz, and the latter is invertible (by SEM.1–2 and SEM.5) it follows by Lemma A.1(v) that $\tilde{m}_0$ is differentiable a.e., with

$$D\tilde{m}_0(w) = D\tilde{r}_0[r_0^{-1}(w)]Dr_0^{-1}(w) = D\tilde{r}_0[r_0^{-1}(w)][Dr_0(w)]^{-1} \tag{A.17}$$

which has nonzero determinant a.e., in view of SEM.2. Thus there exists a set $\mathcal{B} \subset \mathbb{R}^G$, whose complement has measure zero, such that $\tilde{m}_0^{-1}$ is differentiable at the value of $\tilde{m}_0(w)$, for every

$w \in \mathcal{B}$. Taking $w \in \mathcal{B}$, $\lambda > 0$ and $d \in \mathbb{R}^G \backslash \{0\}$, and setting $v = \lambda d$, we obtain from (A.15) that

$$\begin{aligned}
&\lambda^{-1}[\{\tilde{m}_0^{-1}[\lambda d + \tilde{m}_0(w)] + u^* - w\} - u^*] \\
&\qquad = \lambda^{-1}[\{\tilde{m}_0^{-1}[\lambda d + \tilde{m}_0(w)] + u^* - w\} - \{\tilde{m}_0^{-1}[\tilde{m}_0(w)] + u^* - w\}] \\
&\qquad \to (D\tilde{m}_0^{-1})[\tilde{m}_0(w)]d \\
&\qquad = [D\tilde{m}_0(w)]^{-1}d
\end{aligned} \tag{A.18}$$

as $\lambda \to 0$. Hence (A.14), (A.15), (A.16) and (A.18) yield

$$\begin{aligned}
\lambda^{-2}[\delta(v) - \delta(0)] &= \lambda^{-2}[f_U\{\tilde{m}_0^{-1}[\lambda d + \tilde{m}_0(w)] + u^* - w] + w\} - f_U(u^*)] \\
&\to -\tfrac{1}{2} d^{\mathsf{T}}[D\tilde{m}_0(w)^{\mathsf{T}}]^{-1} H [D\tilde{m}_0(w)]^{-1} d \\
&= -\tfrac{1}{2} d^{\mathsf{T}}\{[D\tilde{m}_0(w)] H^{-1} [D\tilde{m}_0(w)]^{\mathsf{T}}\}^{-1} d
\end{aligned}$$

as $\lambda \to 0$, for all $w \in \mathcal{D}$ and $d \in \mathbb{R}^G \backslash \{0\}$.

Since the l.h.s. of the preceding does not depend on w (for any value of $\lambda > 0$), the limit on the r.h.s. cannot either. Therefore, fixing a $w_0 \in \mathcal{D}$ we obtain that

$$[D\tilde{m}_0(w)] H^{-1} [D\tilde{m}_0(w)]^{\mathsf{T}} = [D\tilde{m}_0(w_0)] H^{-1} [D\tilde{m}_0(w_0)]^{\mathsf{T}} =: S$$

for all $w \in \mathcal{D}$. Taking A and B to be the (lower triangular) Cholesky roots of the positive definite matrices $H^{-1} = AA^{\mathsf{T}}$ and $S^{-1} = BB^{\mathsf{T}}$ respectively, it follows that

$$B^{\mathsf{T}}[D\tilde{m}_0(w)] A A^{\mathsf{T}} [D\tilde{m}_0(w)]^{\mathsf{T}} B = B^{\mathsf{T}} S B = I_G$$

for all $w \in \mathcal{D}$, and hence the map

$$\tilde{\ell}_0(w) \coloneqq B^{\mathsf{T}} \tilde{m}_0(Aw)$$

is a locally Lipschitz bijection $\mathbb{R}^G \to \mathbb{R}^G$ for which

$$D\tilde{\ell}_0(w) = B^{\mathsf{T}} D\tilde{m}_0(Aw) A,$$

for all $w \in \mathcal{D}$, and hence

$$\begin{aligned}
D\tilde{\ell}_0(w) D\tilde{\ell}_0(w)^{\mathsf{T}} &= [B^{\mathsf{T}} D\tilde{m}_0(Aw) A][B^{\mathsf{T}} D\tilde{m}_0(Aw) A]^{\mathsf{T}} \\
&= B^{\mathsf{T}} [D\tilde{m}_0(Aw)] A A^{\mathsf{T}} [D\tilde{m}_0(Aw)]^{\mathsf{T}} B = I_G
\end{aligned}$$

for all $w \in \mathcal{D}$, whence also $D\tilde{\ell}_0(w)^{\mathsf{T}} D\tilde{\ell}_0(w) = I_G$ for all $w \in \mathcal{D}$. Moreover, in view of (A.17), SEM.2, and the fact that the determinants of A and B must be strictly positive, as triangular matrices with strictly positive diagonal entries, we have

$$\det D\tilde{\ell}_0(w) = (\det B)[\det D\tilde{m}_0(Aw)](\det A) > 0$$

for all $w \in \mathcal{D}$. Deduce $D\tilde{\ell}_0(w) \in \mathbb{O}^+(G)$ for all $w \in \mathcal{D}$.

It therefore follows by Lemma A.1(v) that there exists a $P \in \mathbb{O}^+(G)$ such that

$$\tilde{\ell}_0(w) = a + Pw.$$

Thus $\tilde{\ell}_0$ is affine, and hence so too is $\tilde{m}_0$.

(v) Conclusion.

To conclude the proof, we recall that $\tilde{m}_0 = \tilde{r}_0 \circ r_0^{-1}$. By the previous part of the proof, there exist $Q \in \mathbb{R}^{G \times G}$ and $q \in \mathbb{R}^G$ such that

$$\tilde{r}_0[r_0^{-1}(w)] = \tilde{m}_0(w) = q + Qw$$

for all $w \in \mathbb{R}^G$, whence taking $y = r_0^{-1}(w)$ yields

$$\tilde{r}_0(y) = q + Qr_0(y)$$

for all $y \in \mathbb{R}^G$.

It similarly follows from (A.13) above that

$$\tilde{r}_1(x) = \tilde{u}^* - \tilde{m}_0[u^* - r_1(x)] = (\tilde{u}^* - q - Qu^*) + Qr_1(x).$$

Hence, defining $q_0 \coloneqq \tilde{u}^* - Qu^*$, we obtain

$$\tilde{U} \coloneqq \tilde{r}_0(Y) + \tilde{r}_0(X) = q_0 + Q[r_0(Y) + r_1(X)] = q_0 + QU$$

whereupon for the distribution of $\tilde{U}$ to respect to scale and location normalisation specified in SEM.3, we must have $q_0 = 0$, and that Q is an orthogonal matrix. Since

$$\det Dr_0(y) = (\det Q)[\det D\tilde{r}_0(y)]$$

a.e., it follows from SEM.2 that $\det Q > 0$, and hence $Q \in \mathbb{O}^+(G)$. □

A.4 Proof of Theorem 2.2

This is essentially a matter of mapping the notation and assumptions imposed on the the nonlinear SVAR in Section 2.2, into their counterparts for the nonlinear SEM in Appendix A.1, and then applying Theorem A.1. Making the identification

$$(Y, X, U) = (z_t, z_{t-1}, \varepsilon_t), \qquad (r_0, r_1, f_U) = (f_0, \boldsymbol{f}_1, \varrho), \qquad (\Gamma_0, \Gamma_1, \Phi) = (\mathscr{F}_0, \boldsymbol{\mathscr{F}}_1, \mathscr{R}), \tag{A.19}$$

and noting that the nonlinear SVAR satisfies PS and DGP, it follows that the nonlinear SEM satisfies SEM, with the only exceptions that: $\det D\tilde{r}_0(y) \neq 0$ a.e., for each $\tilde{r}_0 \in \Gamma_0$ rather than necessarily being strictly positive a.e.; and that the location normalisation $\tilde{r}_0(0) = 0$ is now imposed.

However, since the sign of the determinant of the Jacobian of a locally Lipschitz bijection

$\mathbb{R}^G \to \mathbb{R}^G$ must be the same a.e., it must be the case that for every $\tilde{r}_0 \in \Gamma_0$, either $D\tilde{r}_0(y) > 0$ a.e., or $D\tilde{r}_0(y) < 0$ a.e. Fixing a $Q_0 \in \mathbb{O}(p)$ with $\det Q_0 = -1$, and suppose e.g. that r_0 has $\det Dr_0(y) < 0$ a.e. Then simply by multiplying (A.1) through by Q_0,

$$Q_0 U = (Q_0 r_0)(Y) + (Q_0 r_1)(X),$$

we obtain a parametrisation $(Q_0 r_0, Q_0 r_1, f_{Q_0 U})$ that is observationally equivalent to (r_0, r_1, f_U), but where now $\det D[Q_0 r_0](y) > 0$ a.e. By similarly transforming any candidate $(\tilde{r}_0, \tilde{r}_1, \tilde{f}_U)$ for which $D\tilde{r}_0(y) < 0$ a.e., we can thus reduce the problem to the case where both $Dr_0(y) > 0$ a.e., and $D\tilde{r}_0(y) > 0$ a.e., as is contemplated in Theorem A.1. Because of the possibly intervening transformation by Q_0, that result thus implies that for a given $(\tilde{r}_0, \tilde{r}_1) \in \Gamma_0 \times \Gamma_1$, there exists an $f_{\tilde{U}} \in \Phi$ such that $(\tilde{r}_0, \tilde{r}_1, f_{\tilde{U}})$ is observationally equivalent to (r_0, r_1, f_U), if and only there exists a $Q \in \mathbb{O}(G)$ – which need not now be in $\mathbb{O}^+(G)$ – such that

$$\tilde{r}_0(y) + \tilde{r}_1(x) = Q[r_0(y) + r_1(x)]$$

for all $(y, x) \in \mathbb{R}^G \times \mathbb{R}^K$. Because of the location normalisation $\tilde{r}_0(0) = 0 = r_0(0)$, this is equivalent to

$$\tilde{r}_0(y) = Qr_0(y),\ \forall y \in \mathbb{R}^G \qquad\qquad \tilde{r}_1(x) = Qr_1(x),\ \forall x \in \mathbb{R}^K.$$

Transposing this back to the notation of the SVAR, via (A.19) above, yields the result. □

B Proofs for piecewise affine functions

For the proof of Proposition 3.1, we shall need the following auxiliary result, whose proof is given in Appendix B.2 below. Let the convex hull of a collection of matrices $\{A_i\}_{i=1}^k$ be denoted $\text{co}\{A_i\}_{i=1}^k := \{\sum_{i=1}^k \lambda_i A_i \mid \lambda_i \geq 0,\ \sum_{i=1}^k \lambda_i = 1\}$.

Lemma B.1. *Suppose $f : \mathbb{R}^p \to \mathbb{R}^p$ is a piecewise affine function. Then for every $x', x'' \in \mathbb{R}^p$, there exists a $\Phi \in \text{co}\{\Phi^{(\ell)}\}_{\ell=1}^L$ such that*

$$f(x'') - f(x') = \Phi(x'' - x').$$

B.1 Proof of Proposition 3.1

To simplify the notation, throughout we drop the '0' subscript on f_0 in the statement of the proposition, writing it simply as f.

(i). By either Theorem 1 and 4 in Gouriéroux et al. (1980), which are applicable in the piecewise linear and threshold affine cases respectively, $f : \mathbb{R}^p \to \mathbb{R}^p$ is invertible. Being continuous by assumption, it is therefore a homeomorphism, by Theorem 4.3 in Deimling (1985). Since a piecewise affine function is Lipschitz continuous (Scholtes, 2012, Prop. 2.2.7), it remains only to note that the inverse of an (invertible) piecewise affine function is itself piecewise affine (Scholtes, 2012, Prop. 2.3.1).

(ii). Fix $x', x'' \in \mathbb{R}^p$. We have by Lemma B.1 that for every $u \in \mathbb{R}^p$, there exist non-negative $\{\lambda_\ell(u)\}_{\ell=1}^L$ (which depend also on x', x'') such that $\sum_{\ell=1}^L \lambda_\ell(u) = 1$ and

$$f(x'' + u) - f(x' + u) = \left[\sum_{\ell=1}^L \lambda_\ell(u)\Phi^{(\ell)}\right](x'' - x').$$

Hence

$$\begin{aligned} f_K(x'') - f_K(x') &= \int_{\mathbb{R}^p} [f(x'' + u) - f(x' + u)]K(u)\,\mathrm{d}u \\ &= \sum_{\ell=1}^L \left[\int_{\mathbb{R}^p} \lambda_\ell(u)K(u)\,\mathrm{d}u\right]\Phi^{(\ell)}(x'' - x') =: \left[\sum_{\ell=1}^L \mu_\ell \Phi^{(\ell)}\right](x'' - x'). \end{aligned} \tag{B.1}$$

where $\sum_{\ell=1}^L \mu_\ell = 1$. Since the bracketed matrix on the r.h.s. is an element of $\mathrm{co}\{\Phi^{(\ell)}\}_{\ell=1}^L$, it suffices to show that every matrix in that set is invertible.

We first note the following. Suppose A and B are square matrices, with $\det A > 0$ and $\det B > 0$, and $B - A =: uv^\mathsf{T}$ having rank 1. By Cauchy's formula for the determinant of a rank 1 perturbation,

$$\det B = \det(A + uv^\mathsf{T}) = (\det A)(1 + u^\mathsf{T}A^{-1}v),$$

we must have that $u^\mathsf{T}A^{-1}v > -1$. Therefore for every $\lambda \in [0, 1]$,

$$\det(\lambda A + (1 - \lambda)B) = \det[A + (1 - \lambda)uv^\mathsf{T}] = (\det A)[1 + (1 - \lambda)u^\mathsf{T}A^{-1}v] > 0. \tag{B.2}$$

Now suppose that f is threshold affine. Since f is continuous at the thresholds,

$$\phi^{(\ell-1)} + \Phi^{(\ell-1)}x = \phi^{(\ell)} + \Phi^{(\ell)}x \tag{B.3}$$

for all $x \in \mathbb{R}^p$ such that $a^\mathsf{T}x = \tau_{\ell-1}$. Hence, there exists an $M^{(\ell)} \in \mathbb{R}^{p\times p}$ such that

$$\Phi^{(\ell)} - \Phi^{(\ell-1)} = M^{(\ell)}P_a$$

where $P_a = a(a^\mathsf{T}a)^{-1}a^\mathsf{T}$, whence

$$\Phi^{(\ell)} = \Phi^{(1)} + \sum_{i=2}^{\ell} M^{(i)}P_a =: \Phi^{(1)} + n^{(\ell)}a^\mathsf{T} \tag{B.4}$$

for every $\ell \in \{1, \ldots, L\}$, whereupon the result follows immediately by (B.2), and the fact that $\det \Phi^{(\ell)} > 0$ for all $\ell \in \{1, \ldots, L\}$.

Next suppose that f is piecewise linear, and note that since each $\mathscr{X}^{(\ell)}$ is a union of cones of the form (3.5), we may without loss of generality write

$$f(x) = \sum_{m=1}^{2^p} \mathbf{1}\{x \in \mathscr{C}_{\mathcal{I}_m}\}\tilde{\Phi}^{(m)}x$$

where $\{\mathcal{I}_m\}_{\ell=1}^{2^p}$ partitions $2^{\{1,\ldots,p\}}$, and each for each $m \in \{1, \ldots, 2^p\}$, there is an $\ell \in \{1, \ldots, L\}$

such that $\tilde{\Phi}^{(m)} = \Phi^{(\ell)}$. Moreover, since $A = [a_1, \ldots, a_p]$ is invertible, we may write

$$\begin{aligned} x \in \mathscr{C}_{\mathcal{I}} &\iff a_i^{\mathsf{T}} A^{-1} A x \geq 0, \ \forall i \in \mathcal{I} \text{ and } a_i^{\mathsf{T}} A^{-1} A x < 0, \ \forall i \notin \mathcal{I} \\ &\iff Ax \in \mathscr{D}_{\mathcal{I}} \end{aligned}$$

where

$$\mathscr{D}_{\mathcal{I}} \coloneqq \{x \in \mathbb{R}^p \mid e_i^{\mathsf{T}} x \geq 0, \ \forall i \in \mathcal{I} \text{ and } e_i^{\mathsf{T}} x < 0, \ \forall i \notin \mathcal{I}\}.$$

Hence

$$g(Ax) \coloneqq f[A^{-1}(Ax)] = f(x) = \sum_{m=1}^{2^p} \mathbf{1}\{Ax \in \mathscr{D}_{\mathcal{I}_m}\} \tilde{\Phi}^{(m)} A^{-1}(Ax)$$

and thus it suffices to prove the result with f replaced by

$$g(y) = \sum_{m=1}^{2^p} \mathbf{1}\{y \in \mathscr{D}_{\mathcal{I}_m}\} \Psi^{(m)} y = \sum_{i=1}^{p} [\psi_i^+ \mathbf{1}^+(y) + \psi_i^- \mathbf{1}^-(y)] y_i$$

where $\Psi^{(m)} \coloneqq \tilde{\Phi}^{(m)} A^{-1}$. The second equality holds since g is continuous, and so the coefficient on $y_{i_0} = e_{i_0}^{\mathsf{T}} y$ can only change at the point where $y_{i_0} = 0$; and therefore $\Psi^{(m)} e_{i_0} = \psi_{i_0}^+$ for all m such that $\mathcal{I}_m \ni i_0$, while $\Psi^{(m)} e_{i_0} = \psi_{i_0}^-$ for all m such that $\mathcal{I}_m \not\ni i_0$. By the requirement (3.6), the determinants of each $\Psi^{(m)}$ must have the same sign; we may suppose without loss of generality that these are all positive. Thus it suffices to show that for each $\lambda \coloneqq \{\lambda_m\}_{m=1}^{2^p}$ in the unit simplex,

$$\Psi_\lambda \coloneqq \sum_{m=1}^{2^p} \lambda_m \Psi^{(m)}$$

has $\det \Psi_\lambda > 0$.

To that end, for each $i \in \{1, \ldots, p\}$, define $\mu_i \coloneqq \sum_{m=1}^{2^p} \lambda_m \mathbf{1}\{\mathcal{I}_m \ni i\}$, which sums the weights $\{\lambda_m\}$ over those $\mathscr{D}_{\mathcal{I}_m}$ for which $y_i \geq 0$. Thus the ith column of Ψ_λ is equal to

$$\mu_i \psi_i^+ + (1 - \mu_i) \psi_i^- =: \bar{\psi}_i.$$

For $q \in \{0, \ldots, p\}$, consider the 2^{p-q} matrices defined by

$$\Psi_q(s) = \left[\bar{\psi}_1, \quad \ldots, \quad \bar{\psi}_q, \quad \psi_{q+1}(s_1), \quad \ldots, \quad \psi_p(s_{p-q})\right]$$

where $s \in S^{p-q} \coloneqq \{-1, +1\}^{p-q}$, and

$$\psi_i(u) \coloneqq \psi_i^- \mathbf{1}\{u = -1\} + \psi_i^+ \mathbf{1}\{u = +1\}$$

Then $\Psi_p = \Psi_\lambda$; we will show that this has positive determinant via an induction.

To that end, suppose that $q = 1$. Then for every $s \in S^{p-1}$,

$$\begin{aligned} \Psi_1(s) &= \left[\mu_1 \psi_1^+ + (1 - \mu_1)\psi_1^-, \quad \psi_2(s_2), \quad \ldots, \quad \psi_p(s_p)\right] \\ &= \mu_1 \left[\psi_1^+, \quad \psi_2(s_2), \quad \ldots, \quad \psi_p(s_p)\right] + (1 - \mu_1) \left[\psi_1^-, \quad \psi_2(s_2), \quad \ldots, \quad \psi_p(s_p)\right] \\ &= \mu_1 \Psi_0((+1, s^{\mathsf{T}})^{\mathsf{T}}) + (1 - \mu_1) \Psi_0((-1, s^{\mathsf{T}})^{\mathsf{T}}) \end{aligned}$$

Since both $\Psi_0((+1, s^\mathsf{T})^\mathsf{T})$ and $\Psi_0((-1, s^\mathsf{T})^\mathsf{T})$ are elements of $\{\Psi^{(m)}\}_{m=1}^{2^p}$, they each have positive determinant. Moreover, they differ only by a rank one matrix, and so it follows by (B.2), that $\det \Psi_1(s) > 0$ for all $s \in S^{p-1}$.

Now suppose the claim is true for all $q \in \{1, \ldots, q_0\}$, where $q_0 \leq p-1$. We must show it holds for $q = q_0 + 1$. Consider

$$\begin{aligned}\Psi_{q_0+1}(s) &= \left[\bar{\psi}_1, \quad \ldots, \quad \bar{\psi}_{q_0}, \quad \mu_{q_0+1}\psi^+_{q_0+1} + (1-\mu_{q_0+1})\psi^-_{q_0+1}, \quad \psi_{q_0+1}(s_1), \quad \ldots, \quad \psi_p(s_{p-q_0})\right] \\ &= \mu_{q_0+1}\Psi_{q_0}((+1, s^\mathsf{T})^\mathsf{T}) + (1-\mu_{q_0+1})\Psi_{q_0}((-1, s^\mathsf{T})^\mathsf{T}).\end{aligned}$$

By the inductive hypothesis, both $\Psi_{q_0}((1, s^\mathsf{T})^\mathsf{T})$ and $\Psi_{q_0}((0, s^\mathsf{T})^\mathsf{T})$ have strictly positive determinant; and again they differ only be a rank one matrix. Hence (B.2) implies that $\det \Psi_{q_0+1}(s) > 0$ for all $s \in S^{p-(q_0+1)}$. Deduce that $\Psi_\lambda = \Psi_p$ has strictly positive determinant, and is therefore invertible; deduce that the smoothed counterpart of g, and therefore of f also, is invertible.

We have thus shown that f_K is invertible in both the piecewise linear and threshold affine cases, and that moreover $\sum_{\ell=1}^L \mu_\ell \Phi^{(\ell)}$ in (B.1) has strictly positive determinant. Clearly, f_K is Lipschitz, since $\sum_{\ell=1}^L \mu_\ell \Phi^{(\ell)}$ is bounded, for $(\mu_1, \ldots, \mu_L)$ an element of the $(L-1)$-dimensional unit simplex Δ^{L-1}. It follows moreover that it is bi-Lipschitz, since the final term on the r.h.s. of

$$\|f_K(x'') - f_K(x')\| \geq \|x'' - x'\| \inf_{\|v\|=1} \inf_{\{\mu_\ell\} \in \Delta^{L-1}} \left\| \left[\sum_{\ell=1}^L \mu_\ell \Phi^{(\ell)} \right] v \right\|,$$

may never be zero, since that would otherwise contradict the invertibility of $\sum_{\ell=1}^L \mu_\ell \Phi^{(\ell)}$. By continuity and compactness, the infimum must be achieved, and be strictly positive. Finally, in view of the integrability condition (3.13), f_K must also have m continuous derivatives, by the dominated derivatives theorem. □

B.2 Proof of Lemma B.1

Let $\phi(x) \coloneqq \sum_{\ell=1}^L \mathbf{1}\{x \in \mathscr{X}^{(\ell)}\}\phi^{(\ell)}$ and $\Phi(x) \coloneqq \sum_{\ell=1}^L \mathbf{1}\{x \in \mathscr{X}^{(\ell)}\}\Phi^{(\ell)}$, so that these are constant on each $\mathscr{X}^{(\ell)}$, and $f(x) = \phi(x) + \Phi(x)x$. Now let $x', x'' \in \mathbb{R}^p$; with this notation,

$$\begin{aligned}f(x'') - f(x') &= [\phi(x'') - \phi(x')] + [\Phi(x'')x'' - \Phi(x')x'] \\ &= [\phi(x'') - \phi(x')] + \Phi(x')(x'' - x') + [\Phi(x'') - \Phi(x')]x''. \qquad \text{(B.5)}\end{aligned}$$

Define

$$x(\delta) \coloneqq (1-\delta)x' + \delta x''$$

for $\delta \in [0,1]$. Since f is continuous, so too is $\delta \mapsto f[x(\delta)]$. Because ϕ and Φ are piecewise constant, and $\{\mathscr{X}^{(\ell)}\}_{\ell=1}^L$ is a convex partition of $\mathbb{R}^p$, it follows that $\delta \mapsto \phi[x(\delta)]$ and $\delta \mapsto \Phi[x(\delta)]$ have $m \in \{0, \ldots, L-1\}$ points of discontinuity, located at some $\{\delta_i\}_{i=1}^m$ with $\delta_i < \delta_{i+1}$ for all i. Let $\{x_i\}_{i=1}^{m-1}$ be chosen such that $x_i = x(\delta)$ for some $\delta \in (\delta_i, \delta_{i+1})$, and set $x_0 \coloneqq x'$ and $x_m \coloneqq x''$. By the continuity of $\delta \mapsto f[x(\delta)]$ at each $\delta = \delta_i$, we must have

$$0 = \lim_{\delta \downarrow \delta_i} f[x(\delta)] - \lim_{\delta \uparrow \delta_i} f[x(\delta)] = [\phi(x_i) - \phi(x_{i-1})] + [\Phi(x_i) - \Phi(x_{i-1})]x(\delta_i). \qquad \text{(B.6)}$$

for $i \in \{1, \ldots, m\}$. Noting also that

$$x'' - x(\delta_i) = x'' - [(1-\delta_i)x' + \delta_i x''] = (1-\delta_i)(x'' - x'), \tag{B.7}$$

we may write the final term on the r.h.s. of (B.5) as

$$\begin{aligned}
[\Phi(x'') - \Phi(x')]x'' &= [\Phi(x_m) - \Phi(x_0)]x'' \\
&= \sum_{i=1}^{m} [\Phi(x_i) - \Phi(x_{i-1})]x'' \\
&=_{(1)} \sum_{i=1}^{m} [\Phi(x_i) - \Phi(x_{i-1})][(1-\delta_i)(x'' - x') + x(\delta_i)] \\
&=_{(2)} \sum_{i=1}^{m} (1-\delta_i)[\Phi(x_i) - \Phi(x_{i-1})](x'' - x') - \sum_{i=1}^{m} [\phi(x_i) - \phi(x_{i-1})],
\end{aligned}$$

where $=_{(1)}$ follows from (B.7), and $=_{(2)}$ from (B.6). We note that

$$\sum_{i=1}^{m} [\phi(x_i) - \phi(x_{i-1})] = \phi(x_m) - \phi(x_0)$$

and that setting $\delta_0 \coloneqq 0$ and $\delta_{m+1} = 1$, we have

$$\begin{aligned}
&\sum_{i=1}^{m} (1-\delta_i)[\Phi(x_i) - \Phi(x_{i-1})] \\
&\qquad = (1-\delta_m)\Phi(x_m) + \sum_{i=1}^{m-1} [(1-\delta_i) - (1-\delta_{i+1})]\Phi(x_i) - (1-\delta_1)\Phi(x_0) \\
&\qquad = \sum_{i=0}^{m} [(1-\delta_i) - (1-\delta_{i+1})]\Phi(x_i) - \Phi(x_0) \\
&\qquad = \sum_{i=0}^{m} (\delta_{i+1} - \delta_i)\Phi(x_i) - \Phi(x_0)
\end{aligned}$$

and whence

$$\begin{aligned}
[\Phi(x'') - \Phi(x')]x'' &= \left[\sum_{i=0}^{m} (\delta_{i+1} - \delta_i)\Phi(x_i)\right](x'' - x') - \Phi(x_0)(x'' - x') - [\phi(x_m) - \phi(x_0)] \\
&= \left[\sum_{i=0}^{m} \lambda_i \Phi(x_i)\right](x'' - x') - \Phi(x')(x'' - x') - [\phi(x'') - \phi(x')]
\end{aligned}$$

where $\lambda_i \coloneqq \delta_{i+1} - \delta_i > 0$ and so $\sum_{i=0}^{m} \lambda_i = \sum_{i=0}^{m} (\delta_{i+1} - \delta_i) = \delta_{m+1} - \delta_0 = 1$. It follows from (B.5) that

$$f(x'') - f(x') = \left[\sum_{i=0}^{m} \lambda_i \Phi(x_i)\right](x'' - x').$$

Finally, noting that for each $i \in \{1, \ldots, m\}$, there exists an $\ell_i \in \{1, \ldots, L\}$ such that $\Phi(x_i) = \Phi^{(\ell_i)}$, we have $\sum_{i=0}^{m} \lambda_i \Phi(x_i) \in \mathrm{co}\{\Phi^{(\ell)}\}_{\ell=1}^{L}$ as required. □