Retrospective Estimation of Causal Effects Through Time

Halbert White
University of California, San Diego

Pauline Kennedy
Bates White, LLC

March 28, 2009

Abstract This paper provides methods for estimating a variety of retrospective measures of causal effects in systems of dynamic structural equations. These equations need not be linear or separable. Structural identification of effects of interest is ensured by certain conditional exogeneity conditions, an extension of the notion of strict exogeneity. The covariates ensuring conditional exogeneity can contain not only lags but also leads of suitable proxies for unobservables. We focus on covariate-conditioned average and quantile effects, together with counterfactual objects that are associated with these, such as point bands and path bands. The latter are useful for constructing confidence intervals and testing hypotheses. We show how these objects can be estimated using state-space methods and illustrate with a study of the impact of crude oil prices on gasoline prices.

Acknowledgements We would like to thank Rob Engle, Clive Granger, Jim Hamilton, Òscar Jordà, Xun Lu, Jim Stock, Scott Thompson, and two anonymous referees for their very helpful comments and suggestions. Any errors are the authors' sole responsibility.

1 Introduction

This paper studies methods for retrospectively estimating the causal effects of arbitrary interventions to dynamic economic systems, extending work of White (2006), where the focus was on methods for estimating the effects of natural experiments, e.g., a regime shift.

In pursuing this goal, we blend a number of research themes that have been of central interest to David Hendry throughout his prolific and influential career: *policy analysis* (e.g., Favero and Hendry, 1992; Banerjee, Hendry, and Mizon, 1996; Ericsson, Hendry and Mizon, 1998; Hendry, 2000; Hendry and Mizon, 2000; Hendry, 2002); *dynamic modeling* (e.g., Hendry, 1974; Hendry and Richard, 1982; Hendry, Pagan, and Sargan, 1984; Hendry, 1995c; Hendry, 1996); *forecasting* (e.g., Clements and Hendry, 1996; Clements and Hendry, 1998a,b; Clements and Hendry, 1999; Hendry and Ericsson, 2001; Clements and Hendry, 2002a,b; Clements and Hendry, 2003; Hendry, 2003; Chevillon and Hendry, 2005); *forecast failure* (e.g., Clements and Hendry, 2002a; Hendry, 2002); *notions of exogeneity and their relation to causality* (e.g., Engle, Hendry, and Richard, 1983 (EHR); Engle and Hendry, 1993; Hendry, 1995a; Hendry and Mizon, 1998, 1999; Hendry, 2004); *the links between economics and econometrics* (e.g., Hendry, 1980; Hendry, 1995b; Hendry and Wallace, 1984; Hendry, 1993; Hendry 2001; Hendry, 2005); *cointegration* (e.g., Hendry, 1986; Banerjee and Hendry, 1992); and *automatic modeling* (e.g., Hendry and Krolzig, 2001). Other areas where David Hendry has made seminal contributions, such as *encompassing* (e.g., Hendry and Richard, 1982), are also relevant to our present subject, but are not directly touched on here. These references to Hendry's work are illustrative only. A fuller listing and discussion would leave little space for presenting our own results.

The plan of the paper is as follows. In Section 2, we posit a general dynamic data generating process (DGP), suitable for defining, identifying, and estimating well-defined causal effects. We do not require our dynamic structure to be separable between observable and unobservable variables, nor do we impose other structure, such as linearity or monotonicity. The unobservable drivers of the dependent variable may be countably infinite in dimension. Our framework thus permits analysis of general dynamic treatment effects. These have been considered in depth for panel data in work of Robins (1997) and Abbring and Heckman (2007), among others. In the pure time-series setting considered here, cross-section variation is absent, necessitating the use of methods specific to time-series data.

In Section 3, we define certain *retrospective* covariate-conditioned average effects. Retrospective conditioning makes use of all available relevant information in the past, relative to the present (time T). This creates the novel opportunity to improve predictions for a particular past period ($t < T$) using covariate information from the future relative to that period ($t+\tau, \tau > 0$). Structural identification of effects of interest is ensured by *conditional exogeneity*, a generalization of strict exogeneity, distinct from the notions of weak, strong, and super-exogeneity of EHR, and an extension of White's (2006) CIPP condition.

Section 3 also provides definitions of retrospective covariate-conditioned quantile responses and effects, together with point bands and path bands based on these quantile measures. We discuss how path bands can be used to test hypotheses about the effects of specific interventions.

In Section 4 we propose estimation methods for the effects defined in Section 3. Our estimators belong to a particular class of state-space filters, where counterfactual outcomes central to the definition of the effects of interest play the role of unobservable system states. Section 5 illustrates, with an application to the effects of crude oil prices on gasoline prices. Section 6 contains a summary and concluding remarks.

2 DGP: A Dynamic Structural System

Let y_t denote the values of a $k_y \times 1$ vector (k_y a finite integer) of *responses of interest*, let d_t represent a $k_d \times 1$ vector (k_d a finite integer) of values of response-determining variables whose effects on the response are of primary interest (*causes of interest*), and let v_t and z_t represent countable vectors of values of other response-determining variables (*ancillary causes*). Below we distinguish further between v_t and z_t. We consider a system of structural equations in which the values y_t are generated dynamically as

$$y_t \overset{c}{=} q_t(y^{t-1}, d^t, v^t, z^t), \quad t = 1, 2, ..., \tag{1}$$

where q_t is an unknown $\mathbb{R}^{k_y}$-valued function (the *response function*), $y^{t-1} \equiv (y_0, y_1, ..., y_{t-1})$ denotes the $(t-1)$-*history* of the sequence $\{y_t\}$, and $d^t \equiv (d_0, d_1, ..., d_t)$, $v^t \equiv (v_0, v_1, ..., v_t)$, and $z^t \equiv (z_0, z_1, ..., z_t)$ similarly denote the t-histories of $\{d_t\}, \{v_t\}$, and $\{z_t\}$ respectively.

We enforce the causal direction of time by requiring that only the past and present of the referenced variables determine the time t response. We follow Chalak and White (2007) and White and Chalak (2007a) (WC) in using the notation $\overset{c}{=}$ to emphasize that the structural equations (1) represent directional causal links (Goldberger, 1972, p.979), in which manipulations of elements of y^{t-1}, d^t, v^t, z^t result in differing values for y_t, as in Strotz and Wold (1960) and Fisher (1966, 1972). Leading examples of such structures are those that arise from the dynamic optimization behavior of economic agents and/or interactions among such agents. (See Chow, 1997, for numerous examples.)

Below, we assume that we can observe histories of y_t and d_t, but that we only observe the history of some finitely dimensioned subvectors $\tilde{v}_t$ and $\tilde{z}_t$ of v_t and z_t.

We seek to evaluate certain effects of the causes of interest viewed retrospectively, that is, from the present, time T. Specifying these effects requires special care. Following WC, we define effects in terms of interventions, that is, pairs of alternate values for arguments of the response function. We consider only interventions to the causes of interest. As we take a retrospective view, we focus solely on the effects of *retrospective interventions,* $d^T \to d^{*T} \equiv (d^T, d^{*T})$.

A consequence of the explicit dynamics (lagged y_t's) in eq.(1) is that the effects of interventions can linger, that is, they can propagate through time. To handle this, we can use an alternate implicit dynamic representation. Recursive substitution gives

$$\begin{aligned} y_1 &\overset{c}{=} q_1(y_0, d^1, v^1, z^1) \\ y_2 &\overset{c}{=} q_2(y_0, q_1(y_0, d^1, v^1, z^1), d^2, v^2, z^2) \\ &\vdots \\ y_t &\overset{c}{=} r_t(y_0, d^t, v^t, z^t) \qquad\qquad t = 1, 2, ..., \end{aligned}$$

say, where r_t is an unknown $\mathbb{R}^{k_y}$-valued function that expresses the response value y_t purely in terms of initial values y_0 and the history (d^t, v^t, z^t). We distinguish between q_t and r_t by calling q_t the *explicit* dynamic response function and r_t the *implicit* dynamic response function. With no dynamics, the two are identical.

Analogous to White (2006), we define the *time t ceteris paribus effect of the intervention*

$d^T \to d^{*T}$ *at* (y_0, v^t, z^t) to be

$$\Delta r_t(y_0, d^t, d^{*t}, v^t, z^t) \equiv r_t(y_0, d^{*t}, v^t, z^t) - r_t(y_0, d^t, v^t, z^t), \quad t = 1, ..., T.$$

Because this effect involves the implicit dynamic response function, it fully accounts for any time propagation of effects. Significantly, this effect depends not only on d^t and d^{*t} but also y_0, v^t, and z^t. These values are fixed, consistent with the notion of a *ceteris paribus* effect. Although this defines the effect of $d^T \to d^{*T}$ at time t, only the t-histories d^t, d^{*t}, v^t, and z^t matter, as the elements of d^T, d^{*T}, v^T, and z^T for dates later than t do not determine time t responses.

We ensure that $\Delta r_t(y_0, d^t, d^{*t}, v^t, z^t)$ is the *total effect* of the intervention $d^T \to d^{*T}$ by requiring the system to have the property that y_0, v^t, and z^t do not respond to interventions to d^T. (One can also define and study direct effects and various kinds of indirect effects by imposing other suitable structure (see Chalak and White, 2007, 2008). For conciseness, we focus here solely on total effects.)

Equation (1) does not specify how y_0 is generated; we adopt the convention that y_0 is generated outside the system as the realization of a random vector Y_0. More elaborate conventions are possible. For example, take v_0, z_0 as given, and require that interventions $d^T \to d^{*T}$ satisfy $d_0 = d_0^*$. Or replace d^t, v^t, and z^t in equation (1) with d^{t-1}, v^{t-1}, and z^{t-1}, enforcing a stronger restriction on the operation of causes in time (as advocated by Granger, 1969). Our notation permits flexibility: If contemporaneous effects are allowed, then d_t, v_t, and z_t are observed at time t. If not, then d_t, v_t, and z_t are observed at time $t - 1$. The specifics of any given application often dictate which is more suitable. In any case, we do not permit (d_0, v_0, z_0) to respond to y_0.

For the ancillary causes, we require that v_t and z_t do not vary in response to the histories of d_t or y_t, so that interventions $d^T \to d^{*T}$ have neither direct nor indirect effects on v_t and z_t. If, contrary to this requirement, the dynamic response function is formulated initially in a way that includes ancillary causes that respond to histories of d_t or y_t, one can generally perform substitutions that deliver a system in which this response is absent. Specifically, one can generally express the original "responding" ancillary causes as functions of histories of d_t or y_t and other ancillary causes that do not respond to these histories. With these substitutions, our requirement holds, ensuring that $\Delta r_t(y_0, d^t, d^{*t}, v^t, z^t)$ gives the total effect of the intervention $d^T \to d^{*T}$. When ancillary causes respond to histories of the responses or causes of interest, the system is vulnerable to the Lucas critique (Lucas, 1976). Enforcing the requirement that these dependencies are absent ensures that our system properly captures effects of policy changes represented by interventions.

We enforce these properties by specifying a particular recursive dynamic structure in which "predecessors" structurally determine "successors," but not vice versa. We write $y \Leftarrow d$ to denote that d precedes y (y succeeds d). In particular, future variables (e.g., d_{t+1})

cannot precede present or past variables (e.g., y_t). Necessarily, successors cannot determine predecessors. Predecessors may but do not necessarily cause successors, in the sense defined below. In particular, we specify that

$$\begin{aligned} d_t &\Leftarrow (d^{t-1}, v^t, w^t, z^t) \\ w_t &\Leftarrow (w^{t-1}, v^t, z^t) \\ v_t &\Leftarrow (v^{t-1}, z^t), \qquad\qquad t = 1, 2, ... \end{aligned} \tag{2}$$

Here, we introduce w_t, a finitely dimensioned vector whose t-history may help determine d_t, but not y_t. We thus say that w_t is *structurally irrelevant* for the response of interest. As for y_t and d_t, we observe all elements of w_t. Note that z_t has no predecessors. Thus, we view $\{z_t\}$ as being generated *outside* the system, as the realization of a stochastic process $\{Z_t\}$ with whatever properties may be appropriate for a given application. White and Chalak (2007b) refer to such structurally exogenous variables as *fundamental variables.*

The distinction between v_t and z_t should now be clear: whereas v_t represents ancillary causes determined within the structural system, z_t represents ancillary causes determined outside the structural system.

We formalize the structure developed above as follows:

Assumption A.1 (a) Let $(\Omega, \mathbb{F}, P)$ be a complete probability space, on which are defined random vectors (D_0, V_0, W_0, Y_0) and the stochastic process $\{Z_t\}$, where D_0, V_0, W_0, Y_0, and Z_t take values in $\mathbb{R}^{k_d}$, $\mathbb{R}^{k_v}$, $\mathbb{R}^{k_w}, \mathbb{R}^{k_y}$, and $\mathbb{R}^{k_z}$, respectively, where k_v and k_z are countably valued integers and k_d, k_w, and k_y are finite integers, with $k_d, k_y > 0$, such that $Y_0 \Leftarrow (D_0, V_0, W_0, Z_0)$ and $D_0 \Leftarrow (V_0, W_0, Z_0)$. Further, let $\{D_t, V_t, W_t, Y_t\}$ be a sequence of random vectors such that

$$V_t \Leftarrow (V^{t-1}, Z^t)$$

$$W_t \Leftarrow (W^{t-1}, V^t, Z^t)$$

$$D_t \Leftarrow (D^{t-1}, V^t, W^t, Z^t)$$

$$Y_t \overset{c}{=} q_t(Y^{t-1}, D^t, V^t, Z^t), \qquad t = 1, 2, ...,$$

where q_t is an unknown measurable function taking values in $\mathbb{R}^{k_y}$, and $E(Y_t) < \infty$.

(b) For $t = 0, 1, ...$, $V_t \equiv (\tilde{V}_t, \ddot{V}_t)$ and $Z_t \equiv (\tilde{Z}_t, \ddot{Z}_t)$, where $\tilde{V}_t$ and $\tilde{Z}_t$ take values in $\mathbb{R}^{k_{\tilde{v}}}$ and $\mathbb{R}^{k_{\tilde{z}}}$ respectively, and $k_{\tilde{v}}$ and $k_{\tilde{z}}$ are finite integers. Realizations of $Y_t, D_t, \tilde{V}_t, W_t$, and $\tilde{Z}_t$ are observed; realizations of $\ddot{V}_t$ and $\ddot{Z}_t$ are not observed.

This dynamic structure is quite flexible, as few restrictions are imposed. In particular, we do not require the structural relations to be linear, separable, or monotonic in any of their arguments. Further flexibility can be gained by letting the dimensions of (D_0, V_0, W_0, Y_0) differ from those of (D_t, V_t, W_t, Y_t), $t > 0$. For simplicity, we leave this implicit.

In A.1(a), the referenced measurability refers to measurability-$\mathcal{B}^{\ell_t}/\mathcal{B}^{k_y}$, where $\mathcal{B}^{\ell_t}$ and $\mathcal{B}^{k_y}$ are $\sigma-$fields associated with the domain ($\mathbb{R}^{\ell_t}$) and range ($\mathbb{R}^{k_y}$) of q_t. With k_y (resp. ℓ_t) finite, the $\sigma-$field $\mathcal{B}^{k_y}$ (resp. $\mathcal{B}^{\ell_t}$) is the Borel $\sigma-$field generated by the open sets of $\mathbb{R}^{k_y}$ (resp. $\mathbb{R}^{\ell_t}$). Otherwise, the $\sigma-$field is that generated by the relevant Borel-measurable finite dimensional product cylinders (see, e.g., White, 2001, pp.39-41).

In A.1(b), we specify that V_t and Z_t may not be fully observable. Instead, we observe realizations of finitely dimensioned sub-vectors $\tilde{V}_t$ and $\tilde{Z}_t$, respectively.

The response function q_t contains explicit dynamics. Recursive substitutions give a response with implicit dynamics as

$$Y_t \stackrel{c}{=} r_t(Y_0, D^t, V^t, Z^t), \qquad t = 1, 2, \ldots .$$

The measurability of r_t is ensured by the fact that compositions of measurable functions are again measurable.

3 Defining and Identifying Retrospective Effects

3.1 Average Effects

A key feature of the effect $\Delta r_t(y_0, d^t, d^{*t}, v^t, z^t)$ is that it is empirically inaccessible. That is, we cannot evaluate this effect, even if r_t were known, as not all elements of v_t, z_t are observed. Further, r_t is generally unknown. Nevertheless, it may be possible to estimate useful expected values of the effects of interventions. For this, we introduce some notation. First, let $X_t \equiv (\tilde{V}_t, W_t, \tilde{Z}_t)$ represent the *covariates*; these are observable. We call $U_t \equiv (\ddot{V}_t, \ddot{Z}_t)$ *unobserved causes* and let $U^t \equiv (\ddot{V}^t, \ddot{Z}^t)$ be the t-history of unobserved causes. We write realizations of X^T and U^t as x^T and u^t respectively.

When $E(r_t(Y_0, d^t, V^t, Z^t))$ is finite for each d^t in the support of D^t, we define the *retrospective counterfactual conditional expectation*

$$\begin{aligned} \rho_{t,T}(d^t \mid y_0, x^T) &\equiv E(\ r_t(Y_0, d^t, V^t, Z^t) \mid Y_0 = y_0, X^T = x^T) \\ &= \int r_t(y_0, d^t, v^t, z^t)\ dG_{t,T}(u^t \mid y_0, x^T). \end{aligned}$$

The conditional expectation is "retrospective," as $t \leq T$. We call this expectation "counterfactual" to emphasize that we are *not* conditioning on $D^t = d^t$, as $D^t = d^t$ does not appear in the list of conditioning arguments; we condition only on $(Y_0, X^T) = (y_0, x^T)$. Instead, we view d^t as *set* by some manipulation. The representation of the arguments of $\rho_{t,T}$ is intended to emphasize this distinction. The structure imposed in A.1(a) further ensures that d^t and (y_0, x^T) are variation free: different settings for d^t do not necessitate different values for (y_0, x^T), as (y_0, x^T) is functionally independent of d^t. Thus, $\rho_{t,T}(d^t \mid y_0, x^T)$ gives the expected response conditional on $(Y_0, X^T) = (y_0, x^T)$ for any value of d^t, in particular, for counterfactual values. The integral representation holds under A.1(a), provided

$dG_{t,T}(u^t \mid y_0, x^T)$, the *retrospective conditional density of* U^t *given* $(Y_0, X^T) = (y_0, x^T)$, is regular (Dudley, 2002, ch.10.2). Throughout, we assume that any referenced conditional density is regular.

A noteworthy aspect of $\rho_{t,T}(d^t \mid y_0, x^T)$ is its explicit dependence on "leads" of the covariates, that is, covariate values that occur in the future, relative to the response of interest. For example, if interest attaches to a response at time t, then whenever $t < T$, the expected response $\rho_{t,T}(d^t \mid y_0, x^T)$ can depend on x_{t+1}. Although leads have not received much attention in structural modeling, there is nothing inappropriate about their presence here; indeed, their presence is natural and helpful. Covariate leads do not violate the causal direction of time, as the covariates do not play a causal (structural) role in determining the expected response. They are instead predictive (in the backcasting sense), serving as proxies for unobservable structurally relevant but ancillary causes, U^t. Natural choices for such proxies, as discussed by White (2006), are observed responses $\tilde{V}_t, W_t$, to unobserved ancillary causes U_t of Y_t, observed drivers $\tilde{V}_t, W_t$, $\tilde{Z}_t$ of D_t, and observed responses $\tilde{V}_t, W_t$ to unobserved causes U_t of D_t. (See eq.(2) above.) The presence of dynamics and resultant lingering effects makes it natural that one or more leads of the covariates may be driven by $\ddot{V}_t$ and/or $\ddot{Z}_t$. These leads are thus useful for backcasting U_t.

We use $\rho_{t,T}$ to define the *retrospective covariate-conditioned average effect of intervention* $d^T \to d^{*T}$ as

$$\begin{aligned} \Delta\rho_{t,T}(d^t, d^{*t} \mid y_0, x^T) &\equiv \rho_{t,T}(d^{*t} \mid y_0, x^T) - \rho_{t,T}(d^t \mid y_0, x^T) \\ &= \int \Delta r_t(y_0, d^t, d^{*t}, v^t, z^t)\ dG_{t,T}(u^t \mid y_0, x^T) \\ &= E(\Delta r_t(Y_0, d^t, d^{*t}, V^t, Z^t) \mid Y_0 = y_0, X^T = x^T). \end{aligned}$$

By the optimality property of conditional expectation, we see that $\Delta\rho_{t,T}(d^t, d^{*t} \mid y_0, x^T)$ gives a mean squared error-optimal prediction of $\Delta r_t(Y_0, d^t, d^{*t}, V^t, Z^t)$, the effect of interest, conditional on the specified information $(Y_0 = y_0, X^T = x^T)$. Observe that the ancillary causes v^t, z^t are not held constant here, as they are in $\Delta r_t(y_0, d^t, d^{*t}, v^t, z^t)$. Rather, we *average* over the t-history of unobserved causes U^t, *conditional* on initial values $Y_0 = y_0$ and a T-history of covariates $X^T = x^T$.

There may nevertheless be *ceteris paribus* aspects of $\Delta\rho_{t,T}(d^t, d^{*t} \mid y_0, x^T)$. Specifically, the intervention may hold certain components of the causes of interest constant. For example, if d_t is two dimensional, $d_t = (d_{t1}, d_{t2})$, and we hold d_2^T constant (put $d_2^{*T} = d_2^T$), then $\Delta\rho_{t,T}(d^t, d^{*t} \mid y_0, x^T)$ represents the time t average effect of an intervention to d_1^T $(d_1^T \to d_1^{*T})$ holding d_2^T constant, averaged over the unobserved causes U^t, conditional on the given initial values and the T-history of covariates, y_0, x^T. Besides averages, other aspects of the retrospective conditional distribution of effects can be similarly defined. We discuss some of these in the next subsection.

Although $\rho_{t,T}(d^t \mid y_0, x^T)$ provides the basis for an effect measure whose arguments do not involve unknown quantities, it is nevertheless empirically inaccessible, because it is the conditional expectation of $r_t(Y_0, d^t, V^t, Z^t)$ for counterfactual values d^t, and we have no way to observe $r_t(Y_0, d^t, V^t, Z^t)$. An empirically accessible analog is the *retrospective conditional expectation*

$$\begin{aligned} \mu_{t,T}(y_0, d^t, x^T) &\equiv E(\, Y_t \mid Y_0 = y_0, D^t = d^t, X^T = x^T) \\ &= \int r_t(y_0, d^t, v^t, z^t)\; dG_{t,T}(u^t \mid y_0, d^t, x^T), \end{aligned}$$

where $dG_{t,T}(u^t \mid y_0, d^t, x^T)$ is the *retrospective conditional density of* U^t, *given* $(Y_0, D^t, X^T) = (y_0, d^t, x^T)$, viewing y_0, d^t, and x^T as realizations of random variables Y_0, D^t, and X^T, generated according to Assumption A.1(a). Because this quantity is defined as a functional of the joint distribution of observable variables only, it is empirically accessible, as it can be consistently estimated from a sample of observables under typically mild conditions.

Without further conditions, $\mu_{t,T}$ is purely a stochastic object, providing no information about causal effects. Nevertheless, the equality above shows that the underlying structure embodied in r_t helps determine the properties of $\mu_{t,T}$.

Inspecting $\rho_{t,T}$ and the structural representation for $\mu_{t,T}$, we see that the key difference between them is that $dG_{t,T}(u^t \mid y_0, x^T)$ appears in $\rho_{t,T}$, whereas $dG_{t,T}(u^t \mid y_0, d^t, x^T)$ appears in $\mu_{t,T}$. It follows that if $dG_{t,T}(u^t \mid y_0, d^t, x^T) = dG_{t,T}(u^t \mid y_0, x^T)$ for all u^t, y_0, d^t, and x^T, then $\mu_{t,T} = \rho_{t,T}$. This equality ensures that $\mu_{t,T}$ is not just a stochastic object but also provides structural/causal information. In this case, we say that $\mu_{t,T}$ is *structurally identified*. Similarly, $\rho_{t,T}$ is identified with a stochastic object; we thus say that $\rho_{t,T}$ is *stochastically identified*. When stochastic identification holds uniquely with a representation solely in terms of observable variables, we say that both $\mu_{t,T}$ and $\rho_{t,T}$ are *fully identified* (cf. WC).

The condition $dG_{t,T}(u^t \mid y_0, d^t, x^T) = dG_{t,T}(u^t \mid y_0, x^T)$ for all u^t, y_0, d^t, and x^T is a conditional independence requirement: D^t and U^t are independent given Y_0 and X^T. We express this as

$$D^t \perp U^t \mid Y_0, X^T, \tag{3}$$

following Dawid (1979) (hereafter designated "D"). Because of the similarity to the concept of strict exogeneity (here, $D^t \perp U^t \mid Y_0$) and the central role played by this condition in identifying causal effects, we introduce the following definition.

Definition 3.1 *For given t and $T, t \leq T$, suppose that $D^t \perp U^t \mid Y_0, X^T$. Then we say that D^t is* conditionally exogenous with respect to U^t given (Y_0, X^T).

For brevity, we may just say that D^t is "conditionally exogenous." Conditional exogeneity contains strict exogeneity as the special case in which $X_t \equiv 1$. This concept involves only the DGP and does not involve any parametric model. It is thus distinct from notions of weak, strong, or super-exogeneity of Engle, Hendry, and Richard (1983), as these are defined in terms of the properties of correctly specified parametric models and have primary consequences for estimator efficiency. Conditional exogeneity has no particular implications for estimator efficiency; instead it facilitates identification of causal effects.

The plausibility of conditional exogeneity depends on the structure generating D_t. For example, suppose $D_t \overset{c}{=} c_t(D^{t-1}, X^t, \tilde{U}^t)$, with $(D_0, \tilde{U}^t) \perp U^t \mid Y_0, X^T$, where $\{\tilde{U}_t\}$ is a sequence of unobserved causes of $\{D_t\}$. Then conditional exogeneity holds as a consequence of D, lemmas 4.1 and 4.2.

To proceed, we impose conditional exogeneity.

Assumption A.2 For given T, D^t is conditionally exogenous with respect to U^t given (Y_0, X^T), $t = 1, ..., T$.

Our discussion above establishes the following identification result. For this, we let $\text{supp}(Y_0, X^T)$ denote the support of (Y_0, X^T), that is, the smallest set containing (Y_0, X^T) with probability one, and let $\text{supp}(D^t \mid y_0, x^T)$ denote the support of D^t given that $Y_0 = y_0, X^T = x^T$.

Proposition 3.2 *Suppose Assumptions A.1(a) and A.2 hold. Then* $\mu_{t,T}(y_0, d^t, x^T) = \rho_{t,T}(d^t \mid y_0, x^T)$ *for all* $(y_0, x^T) \in \text{supp}(Y_0, X^T)$ *and* $d^t \in \text{supp}(D^t \mid y_0, x^T)$*. Thus,* $\mu_{t,T}$ *is structurally identified, and* $\rho_{t,T}$ *is stochastically identified. Further,* $\Delta\mu_{t,T} = \Delta\rho_{t,T}$*, so* $\Delta\mu_{t,T}$ *is structurally identified and* $\Delta\rho_{t,T}$ *is stochastically identified, where*

$$\Delta\mu_{t,T}(y_0, d^t, d^{*t}, x^T) \equiv \mu_{t,T}(y_0, d^{*t}, x^T) - \mu_{t,T}(y_0, d^t, x^T).$$

If Assumption A.1(b) also holds, then $\mu_{t,T}$ *and* $\rho_{t,T}$ *are fully identified.*

A measure of effect related to $\Delta\rho_{t,T}$ that is often of interest in applications is

$$y_t - \rho_{t,T}(d^t \mid y_0, x^T),$$

where $y_t = r_t(y_0, d^{*t}, v^t, z^t)$ is the factually observed response value associated with the factual initial value y_0, factual cause histories d^{*t}, v^t, z^t, and factual covariate history x^T; and d^t is a counterfactual scenario of interest. For example, when one is interested in retrospectively measuring the effect of a cartel, y_t represents the price actually generated by the cartel in a particular period t, and d^{*t} is a history representing the operation of the cartel, e.g., $d^*_\tau = 1$ if the cartel operates in period τ and $d^*_\tau = 0$ otherwise. The counterfactual $\rho_{t,T}(d^t \mid y_0, x^T)$ represents the price expected but for the operation of the cartel, under the identical market conditions otherwise. In this case, d^t represents a history in which the cartel did not operate, i.e., a vector of zeroes. We call $y_t - \rho_{t,T}(d^t \mid y_0, x^T)$ a "but-for" average effect. Formally, we have

Corollary 3.3 *Given Assumptions A.1 and A.2,* $y_t - \mu_{t,T}(y_0, d^t, x^T)$ *and* $y_t - \rho_{t,T}(d^t \mid y_0, x^T)$ *are fully identified.*

3.2 Quantile Responses, Quantile Effects, and Point Bands

In applications, effects on aspects of the response distribution other than the average are often of interest. Here we also consider effects defined in terms of *retrospective counterfactual conditional quantiles* for scalar Y_t. (We ensure no loss of generality by letting Y_0 remain a vector.) We begin by defining the *retrospective counterfactual conditional distribution functions*

$$\begin{aligned}\mathcal{F}_{t,T}(y, d^t \mid y_0, x^T) &\equiv P[r_t(Y_0, d^t, V^t, Z^t) \leq y \mid \ Y_0 = y_0, X^T = x^T] \\ &= \int 1[r_t(y_0, d^t, v^t, z^t) \leq y] \ dG_{t,T}(u^t \mid y_0, x^T), \qquad y \in \text{supp}(Y_t \mid y_0, x^T).\end{aligned}$$

The *retrospective counterfactual conditional* $\alpha-$*quantiles* are then given by

$$\mathcal{Q}_{t,T}(\alpha, d^t \mid y_0, x^T) \equiv \inf\{y : \alpha < \mathcal{F}_{t,T}(y, d^t \mid y_0, x^T)\}, \qquad 0 < \alpha < 1.$$

The *retrospective covariate-conditioned* $\alpha-$*quantile effect of intervention* $d^T \to d^{*T}$ is defined by

$$\Delta\mathcal{Q}_{t,T}(\alpha, d^t, d^{*t} \mid y_0, x^T) \equiv \mathcal{Q}_{t,T}(\alpha, d^{*t} \mid y_0, x^T) - \mathcal{Q}_{t,T}(\alpha, d^t \mid y_0, x^T).$$

This measures the impact of the intervention $d^T \to d^{*T}$ on the retrospective conditional $\alpha-$quantile of the response, a version of the covariate-conditioned quantile effect defined by WC, and the unconditional quantile effects of Lehmann (1974), also studied by Abadie, Angrist, and Imbens (2002) and Imbens and Newey (2003).

Observe that $\Delta\mathcal{Q}_{t,T}$ represents a *quantile effect*. The retrospective *effect quantile*, defined as the functional inverse of

$$P[\Delta r_t(Y_0, d^t, d^{*t}, V^t, Z^t) \leq y \mid \ Y_0 = y_0, X^T = x^T],$$

although certainly of interest, is much more complicated to analyze, as it involves the difficult-to-access conditional joint distribution of the responses $r_t(Y_0, d^t, V^t, Z^t)$ and $r_t(Y_0, d^{*t}, V^t, Z^t)$. A detailed consideration of the issues involved for dynamic treatment effect distributions in panel data is given by Abbring and Heckman (2007). Because of the challenges presented by these effects, we leave their consideration aside here. Nevertheless, we consider a related and more tractable measure of effect quantiles below.

The quantile function $\mathcal{Q}_{t,T}$ can be used to define other useful counterfactual objects of interest. For example, for a given time t, the interval

$$[\mathcal{Q}_{t,T}(\alpha/2, d^t \mid y_0, x^T), \quad \mathcal{Q}_{t,T}(1 - \alpha/2, d^t \mid y_0, x^T)]$$

is a *retrospective counterfactual conditional* $(1 - \alpha) \times 100\%$ *confidence interval* for the response under the history d^T. (This is a symmetric interval. Asymmetric intervals can be analogously defined. These may be shorter; we discuss symmetric intervals to keep the notation simple.)

We also call such an interval a $(1-\alpha)$ "point band" for the response at time t, as this interval represents a band of possible responses for a particular point in time having probability $(1-\alpha)$ of containing the true response with history d^T, given $Y_0 = y_0, X^T = x^T$.

Analogous to the but-for average effect measure $y_t - \rho_{t,T}(d^t \mid y_0, x^T)$ introduced above, we can also consider the but-for effect interval

$$[y_t - \mathcal{Q}_{t,T}(1-\alpha/2, d^t \mid y_0, x^T), \quad y_t - \mathcal{Q}_{t,T}(\alpha/2, d^t \mid y_0, x^T)],$$

where y_t is the factually observed response value associated with factual initial value y_0, factual cause histories d^{*t}, v^t, z^t, and factual covariate history x^T; and d^t is a counterfactual scenario whose effects are of interest. In the cartel example discussed above, where y_t represents the time t actual cartel price and d^t represents a history in which the cartel did not operate, this interval represents a range of price outcomes that would have been realized under identical market conditions, but for the operation of the cartel. This range is constructed so as to contain the true but-for outcome with probability $1-\alpha$. In this case, it is possible to measure the effect quantiles, because y_t is set to the actual value.

Stochastic objects analogous to the counterfactual objects just discussed are *retrospective conditional distribution functions*

$$\begin{aligned} F_{t,T}(y \mid y_0, d^t, x^T) &\equiv P[Y_t \leq y \mid \ Y_0 = y_0, D^t = d^t, X^T = x^T] \\ &= \int 1[r_t(y_0, d^t, v^t, z^t) \leq y] \; dG_{t,T}(u^t \mid y_0, d^t, x^T), \qquad y \in \text{supp}(Y_t \mid y_0, d^t, x^T), \end{aligned}$$

and *retrospective conditional $\alpha-$quantiles*,

$$Q_{t,T}(\alpha \mid y_0, d^t, x^T) \equiv \inf\{y : \alpha < F_{t,T}(y \mid y_0, d^t, x^T)\}, \qquad 0 < \alpha < 1.$$

It is now a straightforward exercise to verify

Proposition 3.4 *Suppose Assumptions A.1(a) and A.2 hold. Then* $F_{t,T}(\,\cdot \mid y_0, d^t, x^T) = \mathcal{F}_{t,T}(\,\cdot\,, d^t \mid y_0, x^T)$ *and* $Q_{t,T}(\,\cdot \mid y_0, d^t, x^T) = \mathcal{Q}_{t,T}(\,\cdot\,, d^t \mid y_0, x^T)$ *for all* $(y_0, x^T) \in \text{supp}(Y_0, X^T)$ *and* $d^t \in \text{supp}(D^t \mid y_0, x^T)$, *so* $F_{t,T}$ *and* $Q_{t,T}$ *are structurally identified, and* $\mathcal{F}_{t,T}$ *and* $\mathcal{Q}_{t,T}$ *are stochastically identified.*

Further, $\Delta Q_{t,T} = \Delta \mathcal{Q}_{t,T}$, *so* $\Delta Q_{t,T}$ *is structurally identified, with*

$$\Delta Q_{t,T}(\alpha \mid y_0, d^t, d^{*t}, x^T) \equiv Q_{t,T}(\alpha \mid y_0, d^{*t}, x^T) - Q_{t,T}(\alpha \mid y_0, d^t, x^T),$$

as are the point bands $[Q_{t,T}(\alpha/2 \mid y_0, d^t x^T), \;\; Q_{t,T}(1-\alpha/2 \mid y_0, d^t x^T)]$ *and* $[y_t - Q_{t,T}(1-\alpha/2 \mid y_0, d^t, x^T), y_t - Q_{t,T}(\alpha/2 \mid y_0, d^t, x^T)]$. *In addition, the counterfactual objects* $\Delta \mathcal{Q}_{t,T}$, $[\mathcal{Q}_{t,T}(\alpha/2, d^t \mid y_0, x^T), \;\; \mathcal{Q}_{t,T}(1-\alpha/2, d^t \mid y_0, x^T)]$, *and* $[y_t - \mathcal{Q}_{t,T}(1-\alpha/2, d^t \mid y_0, x^T), y_t - \mathcal{Q}_{t,T}(\alpha/2, d^t \mid y_0, x^T)]$ *are stochastically identified.*

If Assumption A.1(b) also holds, then $F_{t,T}$, $Q_{t,T}$, $\Delta Q_{t,T}$, $\mathcal{F}_{t,T}$, $\mathcal{Q}_{t,T}$, *and* $\Delta \mathcal{Q}_{t,T}$ *are fully identified, as are the corresponding point bands.*

3.3 Path Bands

The point bands introduced above can be used to construct *path bands*, that is sequences of pairs of elements of $\text{supp}(Y_t)$ such that a random sequence of responses, actual or counterfactual, for a given history of causes of interest exits the band defined by the pairs with a specified probability. These path bands are of interest in their own right, as they represent confidence intervals for the response path under the history d^T. Path bands can also be used to test the null hypothesis of the absence of effects for various interventions. They are analogs of the uniform confidence bands for nonparametric regression discussed by Härdle (1990); Jordà and Marcellino (2007) analyze related path bands for prospective forecasts. See also Jordà (2007). To keep the notation simple, we define symmetric path bands; asymmetric path bands can be analogously defined.

Definition 3.5 *A retrospective covariate-conditioned* $(1-\alpha)-$*path band for* $\{r_t(Y_0, d^t, V^t, Z^t)\}$ *over* $[\tau_1, \tau_2]$ *is a sequence of point bands*

$$\{[\mathcal{Q}_{t,T}(\alpha^*/2, d^t \mid y_0, x^T), \quad \mathcal{Q}_{t,T}(1-\alpha^*/2, d^t \mid y_0, x^T)]\}_{t=\tau_1}^{\tau_2}, \tag{4}$$

where α^* *is a function of* α, τ_1, τ_2 *such that*

$$P[\prod_{t=\tau_1}^{\tau_2} 1\{\mathcal{Q}_{t,T}(\alpha^*/2, d^t \mid y_0, x^T) \leq r_t(Y_0, d^t, V^t, Z^t) \leq \mathcal{Q}_{t,T}(1-\alpha^*/2, d^t \mid y_0, x^T)\} \\ = 1 \mid Y_0 = y_0, X^T = x^T] = 1 - \alpha.$$

Although the mathematical expression for this definition is somewhat cumbersome, the basic idea is straightforward: the path band is a sequence of point bands, where the point band coverage, $(1-\alpha^*)$, is chosen so that over the time interval $[\tau_1, \tau_2]$, the path bands contain 100 $(1-\alpha)\%$ of the realized response paths $\{r_t(Y_0, d^t, V^t, Z^t)\}$ generated by d^T, given $Y_0 = y_0$ and $X^T = x^T$. Below, we describe how to estimate $\alpha^* = \alpha^*(\alpha; \tau_1, \tau_2)$.

Equivalently, the (conditional) probability is α that the sequence of responses $\{r_t(Y_0, d^t, V^t, Z^t)\}$ exits the band at any point during the time interval $[\tau_1, \tau_2]$. Thus, one can use the path bands to test the hypothesis

$$H_o : d^T \to d^{*T} \text{ has no effect over the time interval } [\tau_1, \tau_2],$$

where the "effect" is understood to be the effect of the intervention $d^T \to d^{*T}$ on the response of interest, conditional on $Y_0 = y_0$, $X^T = x^T$. To implement the test, one computes the path bands of eq.(4) associated with the "benchmark" history, d^T, for a given level, α. Then one inspects the path of $\{r_t(Y_0, d^{*t}, V^t, Z^t)\}$ to see if it exits the band at any point during the interval $[\tau_1, \tau_2]$. If so, one rejects H_o at level α. Otherwise, one fails to reject H_o. The test relies on the fact that H_o is true if and only if $\{r_t(Y_0, d^{*t}, V^t, Z^t)\}$ has the same conditional distribution (given $Y_0 = y_0$, $X^T = x^T$) on $[\tau_1, \tau_2]$ as $\{r_t(Y_0, d^t, V^t, Z^t)\}$.

In most applications, $\{r_t(Y_0, d^{*t}, V^t, Z^t)\}$ will correspond to an observed response history $\{Y_t\}$ generated by the underlying natural data generating process (subject to the actual history d^{*T}), as the generation of other response histories will require knowledge of the history of the unobserved causes, U^T, as well as knowledge of $\{r_t\}$. Thus, for example, one can test whether a cartel had any effect on prices by comparing the price history generated by the cartel to the path bands generated by the counterfactual history d^T designating the absence of the cartel.

Such tests complement methods of Angrist and Kuersteiner (2004), who propose tests for the presence of causal effects associated with recurring interventions, such as the monetary policy interventions studied by Romer and Romer (1989). The Angrist and Kuersteiner tests make use of the "policy propensity score," rather than directly estimating the effect of the intervention.

4 Estimating Retrospective Effects

When $\rho_{t,T}$ is fully identified, we can estimate $\rho_{t,T}$ by estimating $\mu_{t,T}$. For this, a useful representation of $\mu_{t,T}$ is

$$\mu_{t,T}(y_0, d^t, x^T) = \int y_t \; dF_{t,T}(y_t \mid y_0, d^t, x^T),$$

where $dF_{t,T}(y_t \mid y_0, d^t, x^T)$ defines the conditional density of Y_t given $(Y_0, D^t, X^T) = (y_0, d^t, x^T)$. There are many ways to proceed, but a particularly useful approach is based on estimating $dF_{t,T}(y_t \mid y_0, d^t, x^T)$, as this affords a complete characterization of the conditional stochastic behavior of Y_t. This not only yields estimates of mean effects but also other effects of interest, such as quantile effects or path bands. As $dF_{t,T}$ involves only observable random variables, it can be estimated with suitable data.

Although sample values for Y_t are observable, our interest in counterfactual (thus unobservable) response values under interventions makes it natural to treat the response vector as the state vector for a dynamic state-space system with specific properties appropriate to the present context. This not only permits us to readily develop useful representations for the objects of interest, but also allows us to draw on appropriate segments of the extensive dynamic state-space systems literature.

Viewing Y_t as a state vector, we have the prediction density equation

$$dF_{t+1,T}(y_{t+1} \mid y_0, d^{t+1}, x^T) = \int dF_{t+1,T}(y_{t+1} \mid y_t, y_0, d^{t+1}, x^T) \; dF_{t,T}(y_t \mid y_0, d^{t+1}, x^T). \quad (5)$$

The "filtering" or "updating" density is given by Bayes theorem as

$$dF_{t,T}(y_t \mid y_0, d^{t+1}, x^T) = \frac{dF_{t+1,T}(d_{t+1} \mid y_t, y_0, d^t, x^T) \; dF_{t,T}(y_t \mid y_0, d^t, x^T)}{dF_{t+1,T}(d_{t+1} \mid y_0, d^t, x^T)}. \quad (6)$$

Our assumed DGP permits convenient simplifications. Specifically,

$$dF_{t+1,T}(d_{t+1} \mid y_t, y_0, d^t, x^T) = dF_{t+1,T}(d_{t+1} \mid y_0, d^t, x^T). \tag{7}$$

Equivalently, this condition states that $D_{t+1} \perp Y_t \mid Y_0, D^t, X^T$; this also directly yields

$$dF_{t,T}(y_t \mid y_0, d^{t+1}, x^T) = dF_{t,T}(y_t \mid y_0, d^t, x^T). \tag{8}$$

Formally, we have

Proposition 4.1 *Suppose Assumptions A.1(a) and A.2 hold. Then $D_{t+1} \perp Y_t \mid Y_0, D^t, X^T$, i.e. eqs.(7) and (8) hold.*

Proof: By A.2, $D^{t+1} \perp U^{t+1} \mid Y_0, X^T$. By D lemma 4.2(ii), $D^{t+1} \perp U^{t+1} \mid Y_0, D^t, X^T$. D lemma 4.1 then gives $D^{t+1}, Y_0, X^T \perp U^{t+1}, Y_0, D^t, X^T \mid Y_0, D^t, X^T$. Given A.1(a), we have $Y_t \stackrel{c}{=} r_t(Y_0, D^t, V^t, Z^t)$, so D lemma 4.2(i) gives $D^{t+1}, Y_0, X^T \perp Y_t \mid Y_0, D^t, X^T$. Applying D lemma 4.2(i) once more gives $D_{t+1} \perp Y_t \mid Y_0, D^t, X^T$. ■

The prediction density simplifies under plausible memory restrictions. For concreteness and simplicity, we exploit a "first order" memory condition on the evolution of Y_t,

$$dF_{t,T}(y_t \mid y_{t-1}, y_0, d^t, x^T) = dF_{t,T}(y_t \mid y_{t-1}, d_t, x^T), \tag{9}$$

where the argument lists identify the relevant random variables in the obvious way. Other finite order memory conditions will yield results similar to what follows. We interpret the appearance on the right of d_t in place of d^t as requiring that only D_t has direct predictive relevance for Y_t, given (Y_{t-1}, X^T). Also note that y_0 is absent on the right. Thus, any predictive impact of Y_0 or of past values D^{t-1} is indirect, through Y_{t-1}. (Note that although we permit the effect of D_t to be contemporaneous, this not necessary, as D_t may be observed in period $t-1$, as mentioned above.)

The next result provides a restriction on the dynamic structure sufficient for eq.(9).

Proposition 4.2 *Suppose Assumption A.1(a) holds, that $U_t \perp Y_0 \mid Y_{t-1}, D^t, X^T$ and $U_t \perp D^{t-1} \mid Y_{t-1}, D_t, X^T$ hold, and that*

$$Y_t \stackrel{c}{=} q_t(Y_{t-1}, D_t, V_t, Z_t). \tag{10}$$

Then $Y_t \perp D^{t-1}, Y_0 \mid Y_{t-1}, D_t, X^T$, i.e. eq.(9) holds.

Proof: By D lemma 4.3, $U_t \perp Y_0 \mid Y_{t-1}, D^t, X^T$ and $U_t \perp D^{t-1} \mid Y_{t-1}, D_t, X^T$ imply $U_t \perp Y_0, D^{t-1} \mid Y_{t-1}, D_t, X^T$. Eq.(3.10) and D lemmas 4.1 and 4.2 then give the result. ■

Substituting eq.(9) into eq.(5) gives

$$dF_{t+1,T}(y_{t+1} \mid y_0, d^{t+1}, x^T) = \int dF_{t+1,T}(y_{t+1} \mid y_t, d_{t+1}, x^T)\, dF_{t,T}(y_t \mid y_0, d^{t+1}, x^T). \tag{11}$$

By eq.(8), $dF_{t,T}(y_t \mid y_0, d^{t+1}, x^T) = dF_{t,T}(y_t \mid y_0, d^t, x^T)$. Substituting this into eq.(11) then gives

$$dF_{t+1,T}(y_{t+1} \mid y_0, d^{t+1}, x^T) = \int dF_{t+1,T}(y_{t+1} \mid y_t, d_{t+1}, x^T) dF_{t,T}(y_t \mid y_0, d^t, x^T),$$

which provides a recursion useful for estimating $dF_{t,T}(y_t \mid y_0, d^t, x^T)$. To apply this recursion, we seek an estimate of $dF_{t+1,T}(y_{t+1} \mid y_t, d_{t+1}, x^T)$ for suitable values of t.

Estimation of these densities becomes especially tractable if for given T, there exists a finite non-negative integer τ such that for all $t \leq T - \tau$,

$$dF_{t,T}(y_t \mid y_{t-1}, d_t, x^T) = dF_\tau(y_t \mid y_{t-1}, d_t, x_{t-\tau}^{t+\tau}), \tag{12}$$

where $dF_\tau(y_t \mid y_{t-1}, d_t, x_{t-\tau}^{t+\tau})$ defines the conditional density of Y_t given $Y_{t-1} = y_{t-1}, D_t = d_t$, and $X_{t-\tau}^{t+\tau} = x_{t-\tau}^{t+\tau}$, where $X_{t-\tau}^{t+\tau} \equiv (X_{t-\tau}, ..., X_{t+\tau})$ is the $(\tau-)$ *near history of* X_t. This combines a memory condition with a conditional stationarity assumption. Conditional stationarity holds because dF_τ does not depend on t. The memory condition says that given Y_{t-1} and D_t, only the near history of the covariates is useful in predicting Y_t. This is often plausible, as the memory of U_t contained in the covariates (and the memory of the covariates contained in U_t) will generally fade as time passes. Thus, we impose

Assumption A.3 For given T, there exists a finite non-negative integer τ and a conditional density dF_τ such that for all $t \leq T - \tau$, and for all argument values

$$dF_{t,T}(y_t \mid y_{t-1}, d_t, x^T) = dF_\tau(y_t \mid y_{t-1}, d_t, x_{t-\tau}^{t+\tau}).$$

Combining our results and the above development of the prediction and filtering equations provides the basis for feasible estimation.

Proposition 4.3 *Suppose Assumption A.1(a) holds with eq.(10), and that A.2 and A.3 hold. Then* $dF_{1,T}(y_1 \mid y_0, d^1, x^T) = dF_\tau(y_1 \mid y_0, d_1, x_{1-\tau}^{1+\tau})$ *and for* $t = 1, ..., T - \tau - 1$

$$dF_{t+1,T}(y_{t+1} \mid y_0, d^{t+1}, x^T) = \int dF_\tau(y_{t+1} \mid y_t, d_{t+1}, x_{t+1-\tau}^{t+1+\tau})\ dF_{t,T}(y_t \mid y_0, d^t, x^T).$$

This result makes it straightforward to estimate $dF_{t,T}(y_t \mid y_0, d^t, x^T)$ using sample data when A.1(b) holds. Using that estimate, we can then estimate any desired aspect of the distribution, in particular the means or quantiles that have been our focus here.

From Proposition 4.3, we see that the key to estimating $dF_{t,T}$ is the estimation of dF_τ. Let $d\hat{F}_\tau$ denote any suitable estimator for dF_τ. Depending on the context, one may use either parametric, semi-parametric, or nonparametric estimators $d\hat{F}_\tau$. For example, Li and Racine (2007, ch.5) provide nonparametric methods for conditional density estimation in the empirically relevant "mixed data" case, in which the variables involved may be either continuously or discretely distributed. Given $d\hat{F}_\tau$, we can recursively construct estimators of $dF_{t,T}$ using the structure provided by Proposition 4.3. Specifically, we compute

$$\begin{aligned} d\hat{F}_{1,T}(y_1 \mid y_0, d^1, x^T) &= d\hat{F}_\tau(y_1 \mid y_0, d_1, x_{1-\tau}^{1+\tau}) && (t=0) \\ d\hat{F}_{t+1,T}(y_{t+1} \mid y_0, d^{t+1}, x^T) &= \int d\hat{F}_\tau(y_{t+1} \mid y_t, d_{t+1}, x_{t+1-\tau}^{t+1+\tau})\ d\hat{F}_{t,T}(y_t \mid y_0, d^t, x^T), \\ & \qquad t = 1, ..., T-\tau-1. && (13) \end{aligned}$$

In writing these recursions, we adopt the convention that covariate values for negative time indexes ($t = 1-\tau, ..., -1$) are observable. (Now $X^T \equiv (X_{1-\tau}, ..., X_T)$.) This enables us to maintain our conventions regarding the starting and ending observation indexes for the other variables. It further implies that we can use sample observations $t = 1, ..., T-\tau$ to estimate dF_τ. The recursions above stop τ periods before the end of the sample to accommodate the covariate leads. If it is important to estimate response distributions in periods after $T-\tau$ (e.g., $dF_{T,T}(y_T \mid y_0, d^T, x^T)$), one can modify the procedures above to estimate these.

Using $d\hat{F}_{t,T}$, we estimate $\mu_{t,T}(y_0, d^t, x^T)$ and $\Delta\mu_{t,T}(y_0, d^t, d^{*t}, x^T)$ as

$$\begin{aligned} \hat{\mu}_{t,T}(y_0, d^t, x^T) &= \int y_t\ d\hat{F}_{t,T}(y_t \mid y_0, d^t, x^T) \\ \Delta\hat{\mu}_{t,T}(y_0, d^t, d^{*t}, x^T) &= \hat{\mu}_{t,T}(y_0, d^{*t}, x^T) - \hat{\mu}_{t,T}(y_0, d^t, x^T), \quad t = 1, ..., T-\tau. \end{aligned}$$

Under structural identification, these are also our estimators of $\rho_{t,T}(d^t \mid y_0, x^T)$ and $\Delta\rho_{t,T}(d^t, d^{*t} \mid y_0, x^T)$. The but-for average effect estimator is

$$y_t - \hat{\mu}_{t,T}(y_0, d^t, x^T).$$

To estimate $F_{t,T}$ and $Q_{t,T}$, we can use $d\hat{F}_{t,T}$ to compute

$$\hat{F}_{t,T}(y \mid y_0, d^t, x^T) \equiv \int 1[y_t \leq y]\ d\hat{F}_{t,T}(y_t \mid y_0, d^t, x^T)$$

$$\hat{Q}_{t,T}(\alpha \mid y_0, d^t, x^T) \equiv \inf\{y : \alpha < \hat{F}_{t,T}(y \mid y_0, d^t, x^T)\}, \quad 0 < \alpha < 1.$$

Under structural identification, we can thus estimate $\mathcal{F}_{t,T}$ using $\hat{F}_{t,T}$ and $\mathcal{Q}_{t,T}$ using $\hat{Q}_{t,T}$. The retrospective covariate-conditioned $\alpha-$quantile effect estimator is

$$\Delta\hat{\mathcal{Q}}_{t,T}(\alpha, d^t, d^{*t} \mid y_0, x^T) = \hat{Q}_{t,T}(\alpha \mid y_0, d^{*t}, x^T) - \hat{Q}_{t,T}(\alpha \mid y_0, d^t, x^T).$$

Similarly, the $1-\alpha$ counterfactual point bands can be estimated as

$$[\hat{Q}_{t,T}(\alpha/2 \mid y_0, d^t, x^T),\ \hat{Q}_{t,T}(1-\alpha/2 \mid y_0, d^t, x^T)],$$

and the but-for effect intervals can be estimated as

$$[y_t - \hat{Q}_{t,T}(1-\alpha/2 \mid y_0, d^t, x^T),\ y_t - \hat{Q}_{t,T}(\alpha/2 \mid y_0, d^t, x^T)].$$

To estimate the path bands, it suffices to construct a consistent estimate $\hat{\alpha}^*$ of $\alpha^*(\alpha; \tau_1, \tau_2)$; the path bands are then given by

$$\{[\hat{Q}_{t,T}(\hat{\alpha}^*/2 \mid y_0, d^t, x^T),\ \hat{Q}_{t,T}(1-\hat{\alpha}^*/2 \mid y_0, d^t, x^T)]\}_{t=\tau_1}^{\tau_2}.$$

To construct $\hat{\alpha}^*$, one can use the sequence $\{d\hat{F}_\tau\}$ to generate a large number, say N, of independent and identically distributed (IID) simulated response paths $\{\hat{Y}_{t,i}\}_{t=\tau_1}^{\tau_2}, i = 1, ..., N$, such that for each t and i, $\hat{Y}_{t,i}$ has density $d\hat{F}_\tau(\ \cdot \mid \hat{Y}_{t-1,i}, d_t, x_{t-\tau}^{t+\tau})$. It is then a straightforward numerical exercise to choose $\hat{\alpha}^*$ to solve the problem

$$\min_{\alpha^*} |(1-\alpha) - N^{-1} \sum_{i=1}^{N} \prod_{t=\tau_1}^{\tau_2} 1\{\hat{Q}_{t,T}(\alpha^*/2 \mid y_0, d^t, x^T) \leq \hat{Y}_{t,i} \leq \hat{Q}_{t,T}(1-\alpha^*/2 \mid y_0, d^t, x^T)\}|.$$

Space is not available here to undertake a formal analysis of the properties of these estimators. Because of the close similarity of the estimating equations (13) for $dF_{t,T}$ to those arising in the estimation of state-space models, one may bring the rich array of techniques of that literature to bear in implementing and analyzing the estimators $\{d\hat{F}_{t,T}\}$. Specifically, methods of particle filtering (e.g., Crisan and Doucet, 2002), auxiliary particle filtering (Pitt and Shephard, 1999), or their extensions (e.g., Doucet and Tadić, 2002; Tadić and Doucet, 2002; DeJong, Hariharan, Liesenfeld, and Richard, 2007) are directly relevant.

5 An Illustrative Application

We illustrate the methods described in the previous section by constructing retrospective conditional means, point bands, and path bands useful for examining the impact of crude oil prices on gasoline prices at the monthly frequency. In particular, we study the effects of the Cushing OK WTI spot crude oil price (D_t) on the next month's spot price for U.S. Gulf Coast conventional gasoline (Y_t).

5.1 Gasoline Price Determination

In the present framework, modeling proceeds by identifying the relevant variables of the DGP and then specifying a method for constructing the estimators $\{d\hat{F}_{t,T}\}$. We have already specified Y_t and D_t, so it remains to specify V_t, W_t, and Z_t. As the choice of W_t is primarily informed by that of (V_t, Z_t), we focus first on specifying these variables, the other drivers of gasoline price.

Economic theory says that gasoline prices are determined by the costs of producing gasoline, by demand for gasoline, and by the nature of the conduct among gasoline market participants. For the market and time period we examine (January 1994 through April 2006), we suppose that this conduct is relatively stable. Consequently, we will not include variables to proxy for this conduct. Nevertheless, we can and will use our methods to assess the validity of this assumption. It remains to specify the relevant cost and demand factors.

Crude oil prices are the main driver of gasoline cost, and it is the effect of crude prices on gasoline that is of interest here. To measure the total effect of interest, we thus must omit from consideration cost variables driving gasoline prices that are themselves driven by the crude oil price. This includes such things as crude oil inventories, refining capacity and utilization rates, or diesel fuel prices. Cost factors that may be much less strongly driven by crude oil prices are refinery worker wages, natural gas prices, and interest rates. We treat cost shifters other than crude oil prices as unobservable, belonging to U_t. Thus, we seek proxies for these.

Demand factors not driven by the price of crude oil are regional temperatures and seasonal factors. Income and population are also plausibly weakly driven by crude oil prices in the short run, so we shall treat these also as elements of (V_t, Z_t). Prices of other goods may in principle impact gasoline demand, but for simplicity we assume here that the effects of other prices on gasoline demand are negligible. We thus do not consider these further. We do not assume we can measure the true demand drivers, so we assign these to U_t and seek suitable proxies. Thus, $(\tilde{V}_t, \tilde{Z}_t)$ has zero dimension here.

We also identify drivers of crude oil prices that do not drive gasoline prices and that are not themselves driven by crude oil prices. Such variables are things like exchange rates and industrial production for countries whose growth is not highly dependent on crude oil prices. As for the drivers of gasoline prices, we do not assume these are observable, so we assign them to U_t and seek suitable proxies.

These considerations lead us to select as covariates X_t the following cost and demand proxies W_t: (i) Texas Initial and Continuing Unemployment Claims (taken from State Weekly Claims for Unemployment Insurance Data, Not Seasonally Adjusted); (ii) Houston temperature; (iii) a Winter dummy for January, February, and March; (iv) a Summer dummy for June, July, and August; (v) the 3-Month T-Bill (Secondary Market Rate) (TB3MS); (vi) the U.S. Bureau of Labor Statistics Natural Gas price index; (vii) the U.S. Bureau of Labor Statistics Electricity price index; and (ix) the Yen-US dollar and British pound-US dollar exchange rates.

Our response variable Y_t is the U.S. Gulf Coast Conventional Gasoline, Regular Spot Price (FOB), measured in cents per gallon; D_t is the previous month's Cushing, OK WTI Crude Oil Spot Price (FOB), measured in dollars per barrel.

5.2 Estimation

Figure 1 displays plots of the natural logarithm of gasoline and crude oil prices, together with the change in the natural logarithm of crude oil prices. As expected, gasoline and crude oil prices appear cointegrated. The stochastic trend of oil prices also exhibits an apparent change in January 2002. A test of the null of no change in the mean of the log crude oil price differences before and after January 2002 soundly rejects the null hypothesis of no change. This shift is plausibly thought to be driven by strong growth in demand in East Asia, especially China and India. We examine whether this shift is associated with any corresponding change in the relation between crude oil and gasoline prices. We also examine a counterfactual scenario in which this trend shift is absent.

Accordingly, in a first step, we test for cointegration between these two variables over the period January 1994-December 2001 using the method of Johansen (1991). Finding that these series are cointegrated we estimate a regression model in differences (ΔY_t) by ridge regression. Our regression includes the error correction term and ΔD_t, together with optional lags of the dependent variable and ΔD_t, plus leads and lags of the covariates X_t, transformed to stationarity when appropriate. We explicitly allow changes in crude oil prices to have asymmetric directional impacts. We select variables for the final prediction equation using an automated selection algorithm that implements a general to specific search followed by a specific to general search. At each stage, variables are included or excluded so as to minimize the cross-validated root mean square error (CVRMSE). We also choose the optimal ridge parameter to minimize CVRMSE.

To generate counterfactual retrospective histories, we apply the method of White (2006), in which an initial counterfactual value of d_t is used to generate an initial counterfactual value for Y_t. For successive periods, counterfactual values of d_t are used together with lagged counterfactual values of Y_t to roll forward succeeding counterfactual values of Y_t. In each period, we introduce prediction errors drawn from a normal distribution with standard error equal to the CVRMSE for the estimated prediction equation. This generates a realization of a counterfactual history. Repeating this a large number of times yields conditional means and point bands. We construct path bands from the point bands, as described above. This corresponds to specifying that F_τ is a conditional normal density with conditionally varying mean and conditional homoskedasticity. Other specifications are certainly of interest. We adopt the present specification for simplicity in conducting our illustration.

5.3 Results

First, we construct path bands for the period starting in January 2002 using the actual history of crude oil prices. (Note that these bands are for a period outside the estimation sample.) By comparing actual prices to these path bands, we can test the null hypothesis that there has not been a change in the process generating gasoline prices after 2001 (the "test period"). Among other things, this tests for forecast failure and provides insight into the validity of the market stability assumption introduced above. Figure 2 plots the 5th and 95th percentile path bands around the retrospective dynamic forecast (conditional mean) starting in January 2002. Observed gasoline prices fall within these path bands throughout the entire test period. We thus fail to reject the null hypothesis of stability at the 10% level.

Next, we study crude oil price effects using two alternative counterfactual paths for crude oil prices. Our first counterfactual series is motivated by the apparent structural break in the mean log-difference of crude oil prices. We construct an alternative crude oil series in which no such break occurred by adjusting the post-2001 crude oil price series so that the month-to-month changes in natural logarithms have the same mean as that for the period prior to 2002. Figure 3 shows the actual and resulting counterfactual price series. (The series are constructed using natural log differences and converted to levels for plotting.)

Figure 4 displays 90% path bands for this first counterfactual scenario. Not surprisingly, we see that the actual price exits the path bands, leading to rejection of the null hypothesis of no effect of the change in crude oil price structure on gasoline prices. On average, prices were 54 cents per gallon higher in the period beginning in 2002 than they would otherwise have been, and the gap continues to widen.

Our second counterfactual series is motivated by the disruption to petroleum markets associated with hurricanes Katrina and Rita of 2005. (Katrina reached peak strength on August 28, 2005. We call September 2005 and after the "Katrina period.") We construct an alternative crude oil series representing price behavior plausible in the absence of Katrina and Rita by applying to the Katrina period average month-specific changes for crude-oil price in the periods prior to the hurricanes. Figure 5 displays the actual and resulting counterfactual price series. Figure 6 shows what happens when we estimate the model using data through July, 2005 and then use this to generate counterfactual 90% path bands for the Katrina period. The actual price path begins by spiking up sharply, exiting the counterfactual path bands in September. After the initial price spike, however, the actual price drops to levels below what we would otherwise expect. An expected seasonal price spike for December is absent. Moreover, while gasoline prices are approximately 6 cents above what they would otherwise have been from August 1995 through January 2006, the average impact drops to less than one cent by February 2006, and actual prices are lower than we would expect under our counterfactual scenario.

6 Summary and Conclusion

This paper provides methods for estimating a variety of retrospective measures of causal effects in systems of dynamic structural equations. These equations need not be linear or separable, or possess other properties such as monotonicity. Structural identification of effects of interest is ensured by certain conditional exogeneity conditions, an extension of the notion of strict exogeneity. The variables of the system can be characterized according to their role as responses of interest, causes of interest, ancillary causes, or proxies for unobserved drivers of the responses of interest and the causes of interest. The observed ancillary causes and proxies serve as covariates, playing a supporting role that is predictive rather than causal. Because this predictive role admits back-casting, not only lags but also leads of the covariates may be usefully employed.

We emphasize that only the effects of the causes of interest are informatively measured using our methods. They do not identify effects of observed ancillary causes or the structural dynamics associated with the lags of the dependent variable. Instead, observed ancillary causes and lagged dependent variables form part of the predictive support structure that serves to identify effects of causes of interest.

We pay particular attention to covariate-conditioned average and quantile effects, together with counterfactual objects that are associated with these, such as point bands and path bands. The latter are useful for constructing confidence intervals and testing hypotheses. We show how these objects can be estimated using state-space methods. We illustrate our methods with a study of the impact of crude oil prices on gasoline prices.

There are many interesting topics for further research beyond the scope of this paper. One task is to provide formal conditions ensuring consistency, rates, and/or asymptotic distributions for the estimators proposed here. Another task is to study tests for conditional exogeneity appropriate to the present framework, e.g., extensions of tests proposed by White (2006), so that one can subject hypothesized structures to falsification. The present DGP, with its recursive structure, is only one of a variety of structures in which causal effects can be defined and identified, along the lines of the discussion in Chalak and White (2007); an interesting area for future work is the analysis of identification and estimation of effects in possibly non-recursive systems. Finally, the effects studied here are retrospective; the study of *prospective* effects is also of interest, especially for policy applications. Prospective effects present a variety of interesting analytical challenges distinct from those arising here. Nevertheless, many of the ideas developed here should also prove useful in that study.

References

Abadie, K., Angrist, J., and Imbens, G. (2002), Instrumental variables estimates of the effects of subsidized training on the quantiles of trainee earnings, Econometrica, 70, 91-117.

Abbring, J. and Heckman, J. (2007), Econometric evaluation of social programs, part III: Distributional treatment effects, dynamic treatment effects, dynamic discrete choice, and general equilibrium policy evaluation, in J. Heckman and E. Leamer (Eds.), Handbook of Econometrics, vol 6B. Amsterdam: Elsevier, pp.5145-5303.

Angrist, J. and Kuersteiner, G. (2004), Semiparametric causality tests using the policy propensity score. MIT Department of Economics Working Paper.

Banerjee, A. and Hendry, D.F. (1992), Testing integration and cointegration: An overview, Oxford Bulletin of Economics and Statistics, 54, 225-255.

Banerjee, A., Hendry, D.F., and Mizon, G.E. (1996), The econometric analysis of economic policy, Oxford Bulletin of Economics and Statistics, 58, 573-600.

Chalak, K. and White, H. (2007), An extended class of instrumental variables for the estimation of causal effects. UCSD Department of Economics Discussion Paper.

Chalak, K. and White, H. (2008), Independence and conditional independence in structural systems. UCSD Department of Economics Discussion Paper.

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

Chow, G. (1997). Dynamic Economics: Optimization by the Lagrange Method. Oxford: Oxford University Press.

Clements, M.P. and Hendry, D.F. (1996), Multi-step estimation for forecasting, Oxford Bulletin of Economics and Statistics, 58, 657-684.

Clements, M.P. and Hendry, D.F. (1998a), Forecasting economic processes, International Journal of Forecasting, 14, 111-131.

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

Clements, M.P. and Hendry, D.F. (1999). Forecasting Non-stationary Economic Time Series. Cambridge, MA: MIT Press.

Clements, M.P. and Hendry, D.F. (2002a), Modelling methodology and forecast failure, The Econometrics Journal, 5, 319-344.

Clements, M.P. and Hendry, D.F. (2002b). A Companion to Economic Forecasting. Oxford: Blackwell Publishers.

Clements, M.P. and Hendry, D.F. (2003), Economic forecasting: Some lessons from recent research, Economic Modelling, 20, 301-329.

Crisan, D. and Doucet, A. (2002), A survey of results on particle filtering methods for practitioners, IEEE Transactions on Signal Processing, 50, 736-746.

Dawid, A.P. (1979), Conditional independence in statistical theory, Journal of the Royal Statistical Society, Series B, 41, 1-31.

DeJong, D., Hariharan, D., Liesenfeld, R., and Richard, J.-F. (2007), An efficient filtering approach to likelihood approximation for state-space representations. University of Pittsburgh Department of Economics Discussion Paper.

Doucet, A. and Tadić, V. (2003), Parameter estimation in general state-space models using particle methods, Annals of the Institute of Statistical Mathematics 55, 409-422.

Dudley, R.M. (2002). Real Analysis and Probability. New York: Cambridge University Press.

Engle, R.F. and Hendry, D.F. (1993), Testing super exogeneity and invariance in regression models, Journal of Econometrics, 56, 119-139.

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

Ericsson, N.R., Hendry, D.F. and Mizon, G.E. (1998), Exogeneity, cointegration and economic policy analysis, Journal of Business and Economic Statistics, 16, 1-18.

Favero, C. and Hendry, D.F. (1992), Testing the Lucas critique: A review, Econometric Reviews, 11, 265-306.

Fisher, F. (1966). The Identification Problem in Econometrics. New York: McGraw-Hill.

Fisher, F. (1970), A correspondence principle for simultaneous equations models, Econometrica, 38, 73-92.

Goldberger, A. (1972), Structural equation methods in the social sciences, Econometrica, 40, 979-1001.

Granger, C.W.J. (1969), Investigating causal relations by econometric and cross-spectral methods, Econometrica, 37, 424–438.

Härdle, W. (1990). Applied Nonparametric Regression. Cambridge: Cambridge University Press.

Hendry, D.F. (1974), Stochastic specification in an aggregate demand model of the United Kingdom, Econometrica, 42, 559-578.

Hendry, D.F. (1980). Econometrics: Alchemy or science? Economica, 47, 387-406.

Hendry, D.F. (1986). Econometric modelling with cointegrated variables: An overview, Oxford Bulletin of Economics and Statistics, 48, 201-212.

Hendry, D.F. (1993). Econometrics: Alchemy or Science? Oxford: Oxford University Press.

Hendry, D.F. (1995a), On the interactions of unit roots and exogeneity, Econometric Reviews, 14, 383-419.

Hendry, D.F. (1995b), The role of econometrics in scientific economics, In A. d'Autume and J. Cartelier (Eds.), L'Economie Devient-elle une Science Dure?, pp. 172-196. Paris. Reprinted in English as: Is Economics Becoming a Hard Science? Edward Elgar, 1997.

Hendry, D.F. (1995c). Dynamic Econometrics. Oxford: Oxford University Press.

Hendry, D.F. (1996), Typologies of linear dynamic systems and models, Journal of Statistical Planning and Inference, 49, 177-201.

Hendry, D.F. (2000), Does Money Determine UK Inflation over the Long Run? In R. Backhouse and Salanti, A. (Eds.), Macroeconomics and the Real World. Oxford: Oxford University Press.

Hendry, D.F. (2001). Econometrics: Alchemy or Science? 2nd Edition. Oxford: Oxford University Press.

Hendry, D.F. (2002), Forecast failure, expectations formation, and the Lucas critique, Econometrics of Policy Evaluation, Special Issue of Annales d'Economie et de Statistique, 66-67, pp. 21-40.

Hendry, D.F. (2003), Forecasting pitfalls, Bulletin of EU and US Inflation and Macroeconomic Analysis, 100, 65-82.

Hendry, D.F. (2004), Exogeneity and causality in non-stationary economic processes, in A. Welfe (Ed.), New Directions in Macromodelling, Amsterdam: North Holland, pp. 21–48

Hendry, D.F. (2005), Bridging the gap: Linking economics and econometrics, in C. Diebolt and C. Kyrtsou (Eds.), New Trends in Macroeconomics. Berlin: Springer Verlag.

Hendry, D.F. and Ericsson, N.R. (2001). Understanding Economic Forecasts. Cambridge, MA: MIT Press.

Hendry, D.F. and Krolzig, H.-M. (2001). Automatic Econometric Model Selection. London: Timberlake Consultants Press.

Hendry, D.F. and Mizon, G.E. (1998). Exogeneity, causality, and co-breaking in economic policy analysis of a small econometric model of money in the UK, Empirical Economics, 23, 267–294.

Hendry, D.F. and Mizon, G.E. (1999). The pervasiveness of Granger causality in econometrics, in R.F. Engle and H. White (Eds.), Cointegration, Causality and Forecasting. Oxford: Oxford University Press, pp. 102–134.

Hendry, D.F. and Mizon, G.E. (2000). On selecting policy analysis models by forecast accuracy, in A.B. Atkinson, H. Glennerster, and N.H. Stern (Eds.), Putting Economics to Work: Volume in Honour of Michio Morishima. London School of Economics: STICERD.

Hendry, D.F. and Wallis, K.F. (Eds.) (1984). Econometrics and Quantitative Economics. Oxford: Basil Blackwell.

Hendry, D.F., Pagan, A.R. and Sargan, J.D. (1984), Dynamic specification, in Griliches, Z. and M.D. Intriligator (Eds.), Handbook of Econometrics, Volume 2-3, Chapter 18. Amsterdam: North-Holland.

Hendry, D.F. and Richard, J.-F. (1982). On the formulation of empirical models in dynamic econometrics, Journal of Econometrics, 20, 3-33.

Imbens, G. and Newey, W. (2003), Identification and estimation of triangular simultaneous equations models without additivity. MIT Department of Economics Working Paper.

Johansen, S. (1991), Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models, Econometrica, 59, 1551-1580.

Jordà, Ò. (2007), Simultaneous confidence regions for impulse responses. UC Davis Department of Economics Working Paper.

Jordà, Ò. and Marcellino, M. (2007), Path forecast evaluation. UC Davis Department of Economics Working Paper.

Lehmann, E. (1974). Nonparametrics: Statistical Methods Based on Ranks. San Francisco: Holden-Day.

Li, Q. and Racine, J. (2007). Nonparametric Econometrics: Theory and Practice. Princeton, NJ: Princeton University Press.

Lucas, Robert (1976). Econometric policy evaluation: A critique, Carnegie-Rochester Conference Series on Public Policy, 1, 19–46.

Pitt, M. and Shephard, N. (1999), Filtering via simulation: auxiliary particle filters, Journal of the American Statistical Association, 94, 590-599.

Robins, J. (1997), Causal inference from complex longitudinal data, in M. Berkane (Ed.), Latent Variable Modeling and Applications to Causality. New York: Springer-Verlag, pp.69-117.

Romer, C. and Romer, D. (1989), Does monetary policy matter? NBER Macroeconomics Annual, 121-170.

Strotz, R. and Wold, H. (1960), Recursive vs. nonrecursive systems: An attempt at synthesis, Econometrica, 28, 417-427.

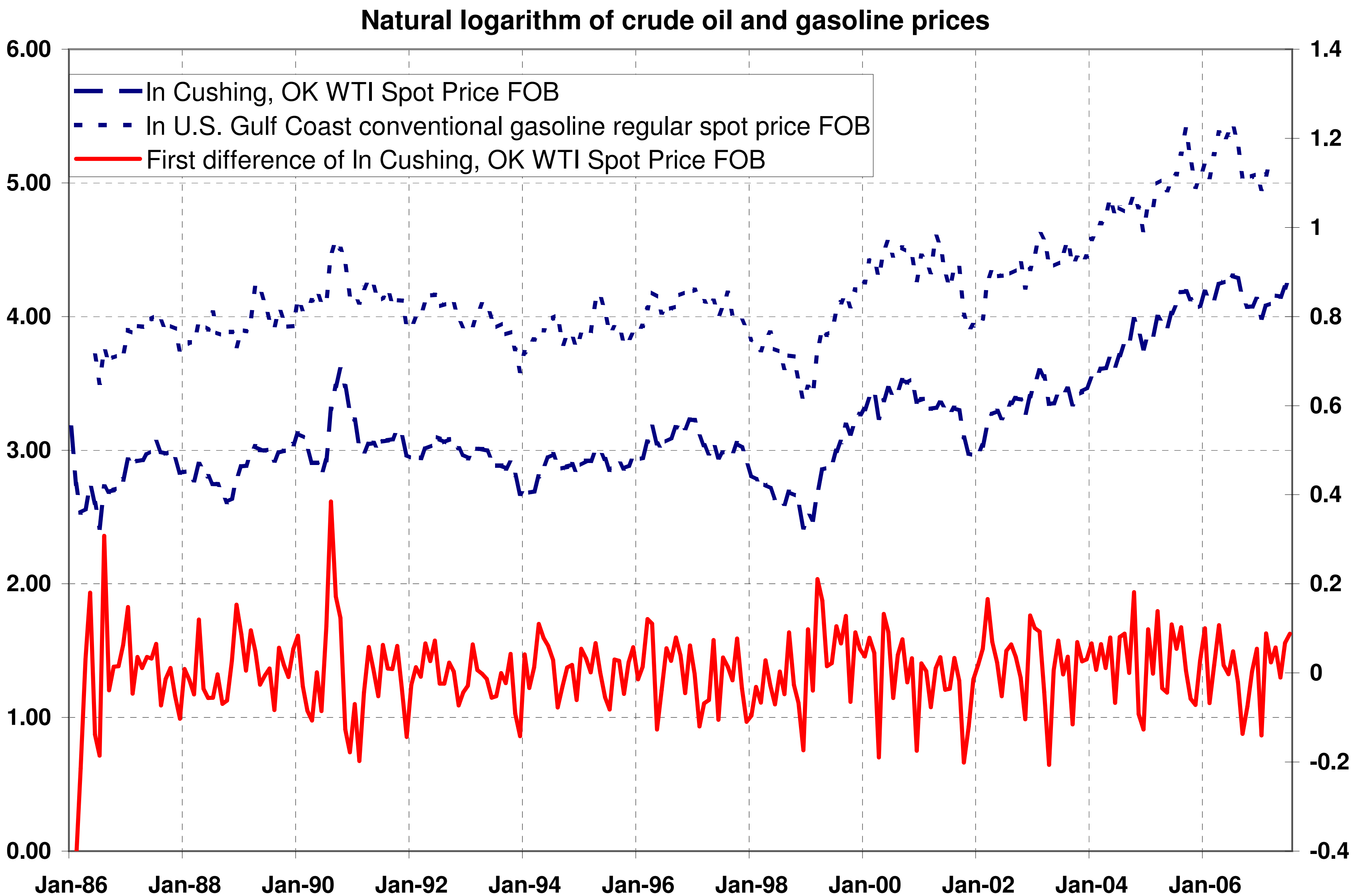
Tadić, V. and Doucet, A. (2002), Exponential forgetting and geometric ergodicity in state-space models, Proceedings of the 41st IEEE Conference on Decision and Control. IEEE Press.

White, H. (2001). Asymptotic Theory for Econometricians. New York: Academic Press.

White, H. (2006), Time-series estimation of the effects of natural experiments, Journal of Econometrics, 135, 527-566.

White, H. and Chalak, K. (2007a), Identifying effects of endogenous causes in nonseparable systems using covariates. UCSD Department of Economics Discussion Paper.

White, H. and Chalak, K. (2007b), Settable systems: An extension of Pearl's causal model with optimization, equilibrium, and learning. UCSD Department of Economics Discussion Paper.



Natural logarithm of crude oil and gasoline prices
6.00
5.00
4.00
3.00
2.00
1.00
0.00
1.4
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
ln Cushing, OK WTI Spot Price FOB
ln U.S. Gulf Coast conventional gasoline regular spot price FOB
First difference of ln Cushing, OK WTI Spot Price FOB
Jan-86
Jan-88
Jan-90
Jan-92
Jan-94
Jan-96
Jan-98
Jan-00
Jan-02
Jan-04
Jan-06

Figure 1

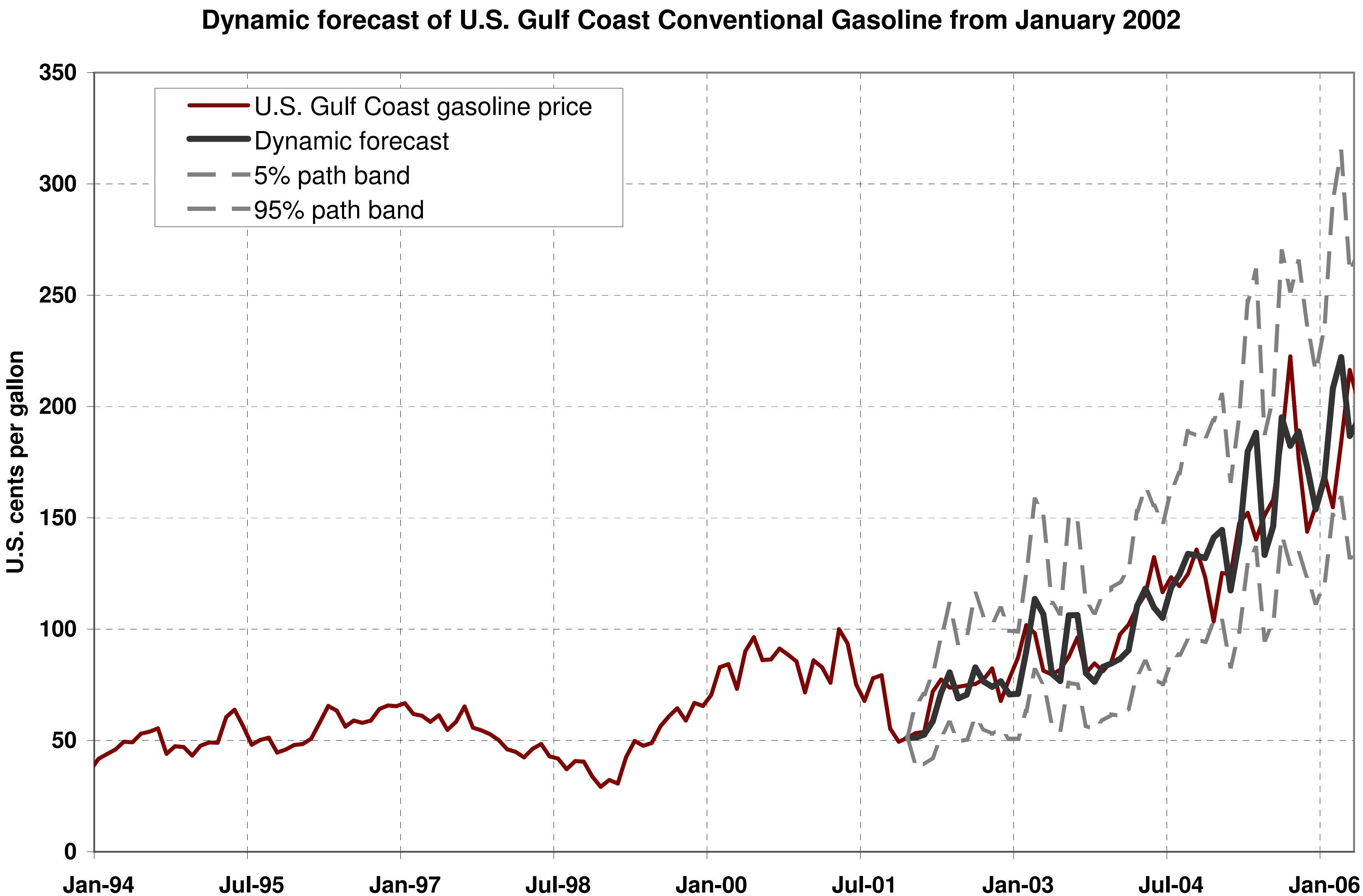

Dynamic forecast of U.S. Gulf Coast Conventional Gasoline from January 2002
U.S. Gulf Coast gasoline price
Dynamic forecast
5% path band
95% path band
U.S. cents per gallon
350
300
250
200
150
100
50
0
Jan-94
Jul-95
Jan-97
Jul-98
Jan-00
Jul-01
Jan-03
Jul-04
Jan-06

Figure 2

Counterfactual of crude oil prices in the absence of a structural break

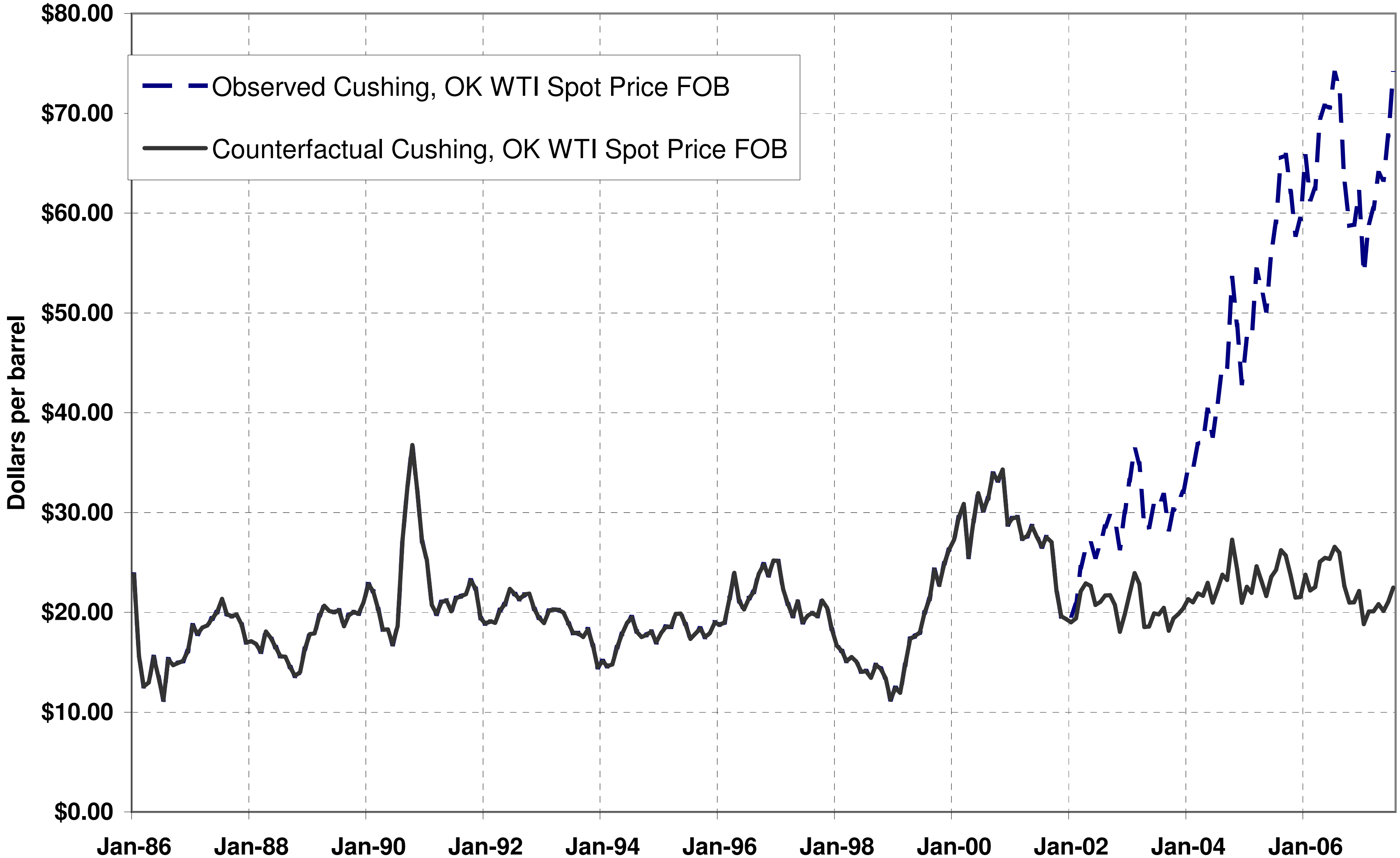


Figure 3

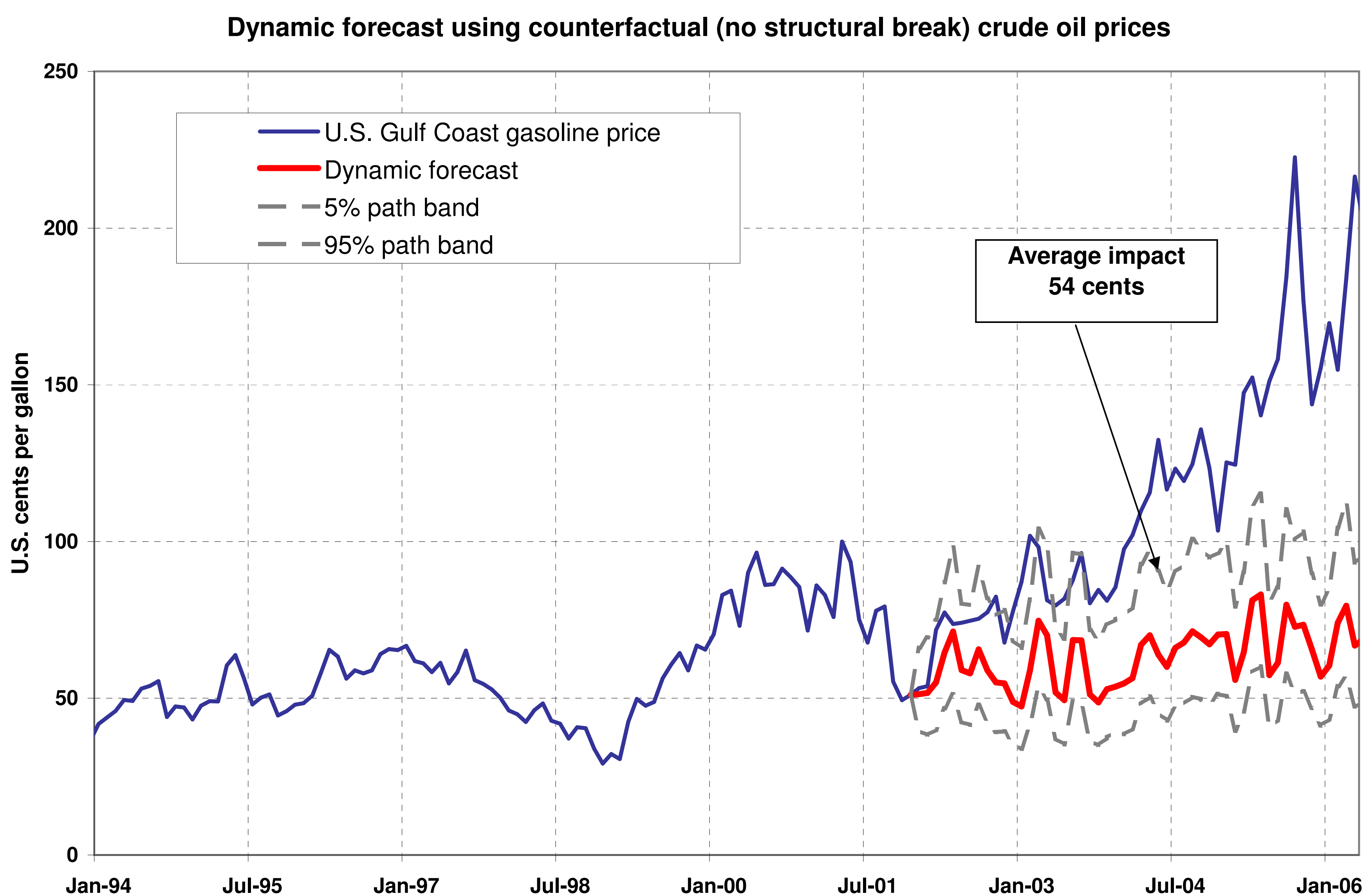

Dynamic forecast using counterfactual (no structural break) crude oil prices
U.S. cents per gallon
250
200
150
100
50
0
U.S. Gulf Coast gasoline price
Dynamic forecast
5% path band
95% path band
Average impact
54 cents
Jan-94
Jul-95
Jan-97
Jul-98
Jan-00
Jul-01
Jan-03
Jul-04
Jan-06

Figure 4

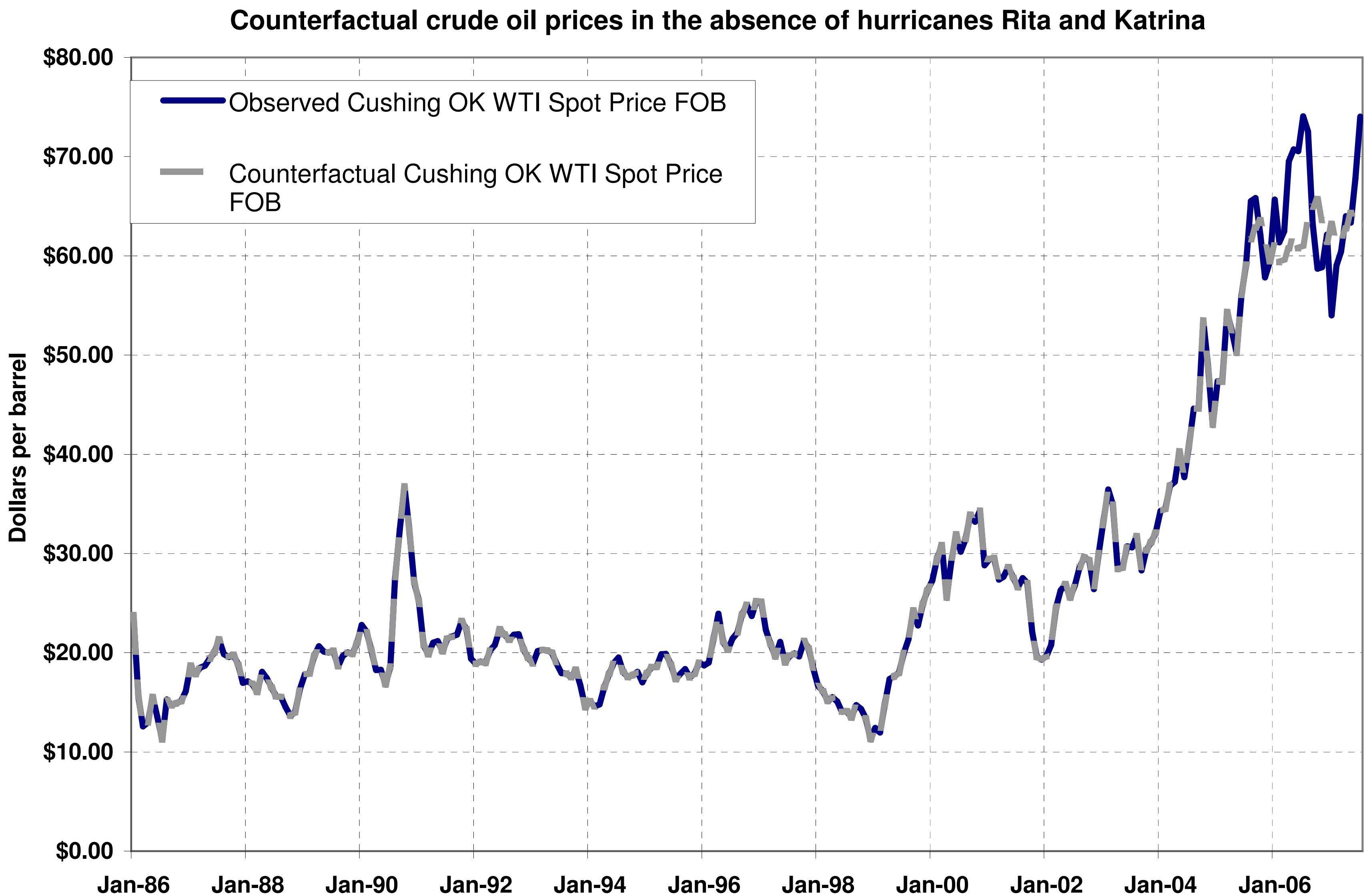

Counterfactual crude oil prices in the absence of hurricanes Rita and Katrina
$80.00
$70.00
$60.00
$50.00
$40.00
$30.00
$20.00
$10.00
$0.00
Dollars per barrel
Observed Cushing OK WTI Spot Price FOB
Counterfactual Cushing OK WTI Spot Price FOB
Jan-86
Jan-88
Jan-90
Jan-92
Jan-94
Jan-96
Jan-98
Jan-00
Jan-02
Jan-04
Jan-06

Figure 5

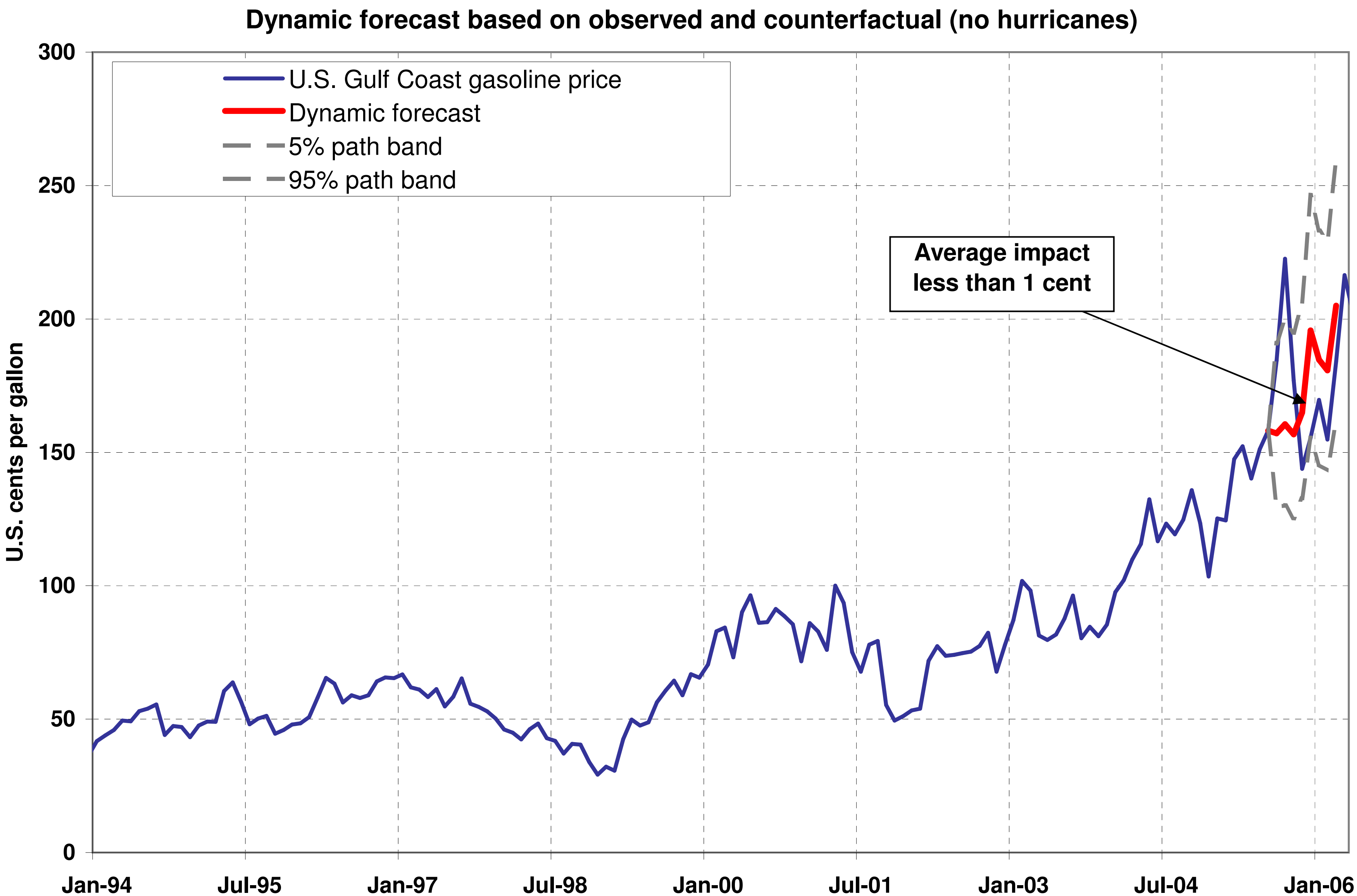

Dynamic forecast based on observed and counterfactual (no hurricanes)
U.S. Gulf Coast gasoline price
Dynamic forecast
5% path band
95% path band
Average impact less than 1 cent
U.S. cents per gallon
300
250
200
150
100
50
0
Jan-94
Jul-95
Jan-97
Jul-98
Jan-00
Jul-01
Jan-03
Jul-04
Jan-06

Figure 6