B Web Appendix for "A nonparametric dynamic causal model for macroeconometrics": additional results

B.1 Information sets and non-anticipation

The key for potential outcome time series is the non-anticipation of potential outcome $Y_t(w_{1:t})$ and treatments, $\Pr(W_t|W_{1:t-1} = w_{1:t-1}, Y_{1:T}(\bullet)) = \Pr(W_t|W_{1:t-1}, Y_{1:t-1}(W_{1:t-1}))$ stated in Assumptions 1 and 2.

Suppose that the treatments are generated with the aid of additional information, $X_{1:T}(\bullet)$. If $X_{1:T} = X_{1:T}(W_{1:T})$ is observed by the econometrician this raises no issues as we can simply augment the potential outcomes time series to include $X_{1:T}$. More precisely, we state this in the following modified non-anticipation assumption.

Assumption 5 (Extended non-anticipation)**.** *For each $t \geq 1$, the pair $\{Y_t(w_{1:T}), X_t(w_{1:T})\}$ satisfies*

$$\{Y_t(w_{1:T}), X_t(w_{1:T})\} = \{Y_t(w_{1:t}), X_t(w_{1:t})\}, \quad \text{and}$$
$$\Pr(W_t|W_{1:t-1}, Y_{1:T}, X_{1:T}) = \Pr(W_t|W_{1:t-1}, Y_{1:t-1}(W_{1:t-1}), X_{1:t-1}(W_{1:t-1})).$$

We observe $\{W_t, Y_t(W_{1:t}), X_t(W_{1:t})\}$. If Assumption 5 holds, then the additional series $X_t(W_{1:t})$ raises no new issues, we can use our methods to learn about the causal effect of W_{t-p} on Y_t or W_{t-p} on X_t.

Example 2. *Return to Example 1 but now allow the optimizing decision be:*

$$W_t = \arg\max_{w_t} \mathbb{E}[\max_{w_{t+1:T}} U(Y_{t:T}(w_{1:T}), w_{t:T}) \,|\, Y_{1:t-1}(W_{1:t-1}), X_{1:t-1}(W_{1:t-1}), W_{1:t-1} = w_{1:t-1}], \quad (22)$$

where U is a utility function of future outcomes and treatments. The expectation is over the law of $Y_{t:T}(\bullet)$. This decision rule delivers W_t and therefore also delivers $\{Y_t(W_{1:t}), X_t(W_{1:t})\}$.

A distinct case is where $X_{1:T}(W_{1:T})$ influences the decision maker but is unobserved seen by the econometrician. This is a natural concern. For example, in estimating the causal effects of monetary policy, we may worry that the Federal Reserve has access to private information about future macroeconomic conditions that is unavailable to the econometrician. To deal with this omitted variables problem, we introduce a stronger assumption.

Assumption 6 (Non-anticipating missing data)**.** *Suppose that Assumption 5 holds for each $t \geq 1$.*

Additionally assume that $X_t(\bullet)$ satisfies, for all $w_{1:t} \in \mathcal{W}^t$ and x_t,

$$\Pr(X_t(w_{1:t}) \leq x_t | X_{1:t-1}(w_{1:t-1}), Y_{1:T}(w_{1:T})) = \Pr(X_t(w_{1:t}) \leq x_t | X_{1:t-1}(w_{1:t-1}), Y_{1:t}(w_{1:t})).$$

This says that the potential missing data is only influenced by past potential missing data and past and current potential outcomes. Hence $Y_{t+1:T}(w_{1:T})$ does not Granger cause $X_t(w_{1:t})$ conditional on past observed and unobserved data. This is a strong assumption and its plausibility is context dependent – it imposes that private and public information is influencing treatments and then treatments and public information impact outcomes. Thus the omitted variable does not have its own causal channel onto outcomes.

If Assumption 6 holds, we are able to show the following important methodological result.

Proposition B.1. *Under Assumption 6,* $\Pr(W_t \leq w_t | W_{1:t-1}, Y_{1:T}(\bullet)) = \Pr(W_t \leq w_t | W_{1:t-1}, Y_{1:t-1}(\bullet))$.

Proof. First note that

$$\Pr(X_{1:T}(w_{1:T}) \leq x_{1:T} | Y_{1:T}(w_{1:T})) = \prod_{t=1}^{T} \Pr(X_t(w_{1:T}) \leq x_t | X_{1:t-1}(w_{1:T}), Y_{1:T}(w_{1:T})),$$

by the prediction decomposition and so

$$\Pr(X_{1:T}(w_{1:T}) \leq x_{1:T} | Y_{1:T}(w_{1:T})) = \prod_{t=1}^{T} \Pr(X_t(w_{1:t}) \leq x_t | X_{1:t-1}(w_{1:t}), Y_{1:t}(w_{1:t}))$$

by Assumption 6. Then

$$\begin{aligned}
\Pr(W_t \leq w_t | W_{1:t-1}, Y_{1:T}(\bullet)) &\overset{(1)}{=} \mathbb{E}^{X_{1:t-1}(\bullet)|W_{1:t-1}, Y_{1:T}(\bullet)}\{\Pr(W_t \leq w_t | W_{1:t-1}, Y_{1:T}(\bullet), X_{1:t-1}(\bullet))\} \\
&\overset{(2)}{=} \mathbb{E}^{X_{1:t-1}(\bullet)|W_{1:t-1}, Y_{1:T}(\bullet)}\{\Pr(W_t \leq w_t | W_{1:t-1}, Y_{1:t-1}(\bullet), X_{1:t-1}(\bullet))\} \\
&\overset{(3)}{=} \mathbb{E}^{X_{1:t-1}(\bullet)|W_{1:t-1}, Y_{1:t-1}(\bullet)}\{\Pr(W_t \leq w_t | W_{1:t-1}, Y_{1:t-1}(\bullet), X_{1:t-1}(\bullet))\}
\end{aligned}$$

where (1) holds by iterated expectations, (2) holds by non-anticipating treatments, (3) holds by non-anticipating missing data. Then the result follows by iterated expectations. □ □

Proposition B.1 implies that if $\{W_t, Y_t(\bullet), X_t(\bullet)\}$ is a potential outcome time series, and $X_t(\bullet)$ obeys Assumption 6, then $\{W_t, Y_t(\bullet)\}$ is a potential outcome time series. Hence the causal methods for W_{t-p} on Y_t we developed earlier still apply.

B.2 Section 2.2: general weight functions to weighted causal effects

Example 1 (continued)**.** *Interpret W_t to be a scaled change in the Federal funds rate. The possible changes are $\mathcal{W} = \{-1, 0, 1\}$ and $w = 1, w' = 0$. There is a unique $\tau_{k,j,t}(1,0)(0)$, which is the contemporaneous causal effect of the change in the Federal funds rate on the unemployment rate. However, when $p = 1$ and $\tau_{k,j,t}(1,0)(1)$, there are 9 possible causal effects corresponding to $w_{t-1:t}$ equalling $(1,-1)$, $(1,0)$ or $(1,-1)$ and $w'_{t-1:t}$ equalling $(0,-1)$, $(0,0)$ or $(0,-1)$. Each examines the impact on the unemployment rate of a particular path of the Federal funds rate from time $t-1$ to time t and may be interesting in their own right. Our approach is to report a weighted version of these 9 different causal effects. A simple choice would be to compare $(1,1)$ with $(0,0)$ and $(1,0)$ with $(0,-1)$. This gives the average causal effect of a rate increase at time $t-1$ and time t.*

Generally, let $G_{t-p-1}(x, y|w, w')$ be some conditional probability measure in x, y for scalars $w, w' \in \mathcal{W}$ and in the potential outcomes, fix $w_{j,t-p} = w$ and $w'_{j,t-p} = w'$. Then, the *general weighted causal effect* is

$$\int_{w_{-j,t-p:t}, w'_{-j,t-p:t}} \{Y_{k,t}(w^{obs}_{1:t-p-1}, w_{t-p:t}) - Y_{k,t}(w^{obs}_{1:t-p-1}, w'_{t-p:t})\}\, dG_{t-p-1}(w_{-j,t-p:t}, w'_{-j,t-p:t}|w, w'), \tag{23}$$

where $w_{-j,t-p:t} = (w_{1:j-1,t-p}, w_{j+1:n_w,t-p}, w_{t-p+1:t})$ denotes all elements of $w_{t-p:t}$ except $w_{j,t-p}$. The general weighted causal effect is a valid causal effect whatever choice of $G_{t-p-1}(x, y|w, w')$ is made as it is a weighted average of time-t causal effects.

In this paper, we set $G_{t-p-1}(x, y|w, w') = G_{t-p-1}(x|w)G_{t-p-1}(y|w')$ for a common conditional probability measure $G_{t-p-1}(\cdot|\cdot)$. That is, the function $G_{t-p-1}(x|w)$ specifies the function $G_{t-p-1}(y|w')$ as well as the function $G_{t-p-1}(x, y|w, w')$. With this restriction, the weighted causal effect can be decomposed into

$$\int \{Y_{k,t}(w^{obs}_{1:t-p-1}, w_{t-p:t})\} dG_t(w_{-j,t-p:t}|w) - \int \{Y_{k,t}(w^{obs}_{1:t-p-1}, w'_{t-p:t})\} dG_t(w'_{-j,t-p:t}|w').$$

This is the difference between the average of the potential outcomes along treatment and counterfactual path respectively. We focus on one particular choice of the weights $G_{t-p-1}(x|w)$ given in Definition 7.

B.3 Additional results for section 3: Nonparametric estimator of the CRF

We can extend the results in Theorem 3.1 to provide an unbiased estimator of the causal response function introduced in Definition 8. We state this in the following proposition.

Proposition B.2. *Assume a potential outcome time series and that Assumption 4 holds. Then,*

$$\mathbb{E}^{W,Y(\bullet)}[\hat{\bar{\tau}}_{k,j,t}(w,w')(p) \,|\, y^{obs}_{1:t-p-1}, w^{obs}_{1:t-p-1}] = CRF_{k,j,t}(w,w')(p)$$

Let $v_{k,j,t-p}(w,w')(p) \equiv \hat{\bar{\tau}}_{k,j,t}(w,w')(p) - CRF_{k,j,t}(w,w')(p)$. *This is a martingale difference sequence with respect to the history of outcomes and treatments up to time* $t-p-1$, $\mathcal{F}_{t-p-1}$ *and the conditional variance, if it exists, is*

$$\zeta^2_{k,j,t-p}(w,w')(p) \equiv Var^{W,Y(\bullet)}(v_{k,j,t-p}(w,w')(p) \,|\, \mathcal{F}_{t-p-1}) = \mathbb{E}^{Y(\bullet)}[\eta^2_{k,j,t-p}(w,w')(p) \,|\, \mathcal{F}_{t-p-1}]$$

Proof. Theorem 3.1 implies that, if it exists,

$$\mathbb{E}^{W,Y(\bullet)}\{\hat{\bar{\tau}}_{k,j,t}(w,w')(p)|y^{obs}_{1:t-p-1}, w^{obs}_{1:t-p-1}\} = \mathbb{E}^{Y(\bullet)}\left(\mathbb{E}^{W|Y(\bullet)}\{\hat{\bar{\tau}}_{k,j,t}(w,w')(p)|w^{obs}_{1:t-p-1}\}|y^{obs}_{1:t-p-1}\right)$$
$$= \mathbb{E}^{Y(\bullet)}(\bar{\tau}_{k,j,t}(w,w')(p)|y^{obs}_{1:t-p-1}) = CRF_{k,j,t}(w,w')(p).$$

This gives the first result. Next, the conditional variance result follows as $v_{k,j,t-p}(w,w')(p)$ is a martingale difference sequence with respect to $\mathcal{F}_{t-p-1}$. □ □

Recall the average weighted causal effect $\bar{\tau}_{k,j}(w,w')(p) := \frac{1}{T-p}\sum_{t=p+1}^{T} \bar{\tau}_{k,j,t}(w,w')(p)$ introduced in Section 2. The average of the Horvitz and Thompson (1952) style estimator

$$\hat{\bar{\tau}}_{k,j}(w,w')(p) \equiv \frac{1}{T-p}\sum_{t=p+1}^{T} \hat{\bar{\tau}}_{k,j,t}(w,w')(p)$$

is an unbiased estimator of the temporal average weighted causal effect over the treatment path holding the potential outcomes fixed. This will not require any additional restrictions on the weighted causal effects – that is, $\bar{\tau}_{k,j,t}(w,w')(p)$ can vary across periods. We state this result in the following proposition.

Proposition B.3. *Assume a potential outcome time series and that Assumption 4 holds. Then,*

$$\mathbb{E}^{W\,|\,Y(\bullet)}[\hat{\bar{\tau}}_{k,j}(w,w')(p)] = \bar{\tau}_{k,j}(w,w')(p).$$

Moreover, if the moments are such the non-anticipating treatment paths satisfy $\lim_{T\to\infty} \frac{1}{T-p}\sum_{t=p+1}^{T} \mathbb{E}^{W\,|\,Y(\bullet)}[\eta^2_{k,j,t-p}(w,w')(p)] < \infty$, *then* $\hat{\bar{\tau}}_{k,j}(p) - \bar{\tau}_{k,j}(p) \xrightarrow{p} 0$ *as* $T \to \infty$.

Proof. The first result follows from an analogous argument as in Theorem 3.1. Next, let $u_{k,j}(w,w')(p) = \frac{1}{T-p}\sum_{t=p+1}^{T} u_{k,j,t-p}(w,w')(p)$ and then, because $u_{k,j,t-p}(w,w')(p)$ is a martingale difference sequence from Theorem 3.1,

$$\bar{\eta}_{k,j}(w,w')(p) = Var^{W|Y(\bullet)}\{\sqrt{T}u_{k,j}(w,w')(p)\} = \frac{T}{T-p}\frac{1}{T-p}\sum_{t=p+1}^{T} \mathbb{E}^{W|Y(\bullet)}[\eta^2_{k,j,t-p}(w,w')(p)].$$

So, under the assumptions of the proposition, as $T \to \infty$, $\sqrt{T}u_{k,j}(w,w')(p) = O_p(1)$. Since $\hat{\bar{\tau}}_{k,j}(w,w')(p) - \bar{\tau}_{k,j}(w,w')(p) = u_{k,j}(w,w')(p)$, the second result follows immediately. □ □

Next, recall the average causal response function defined as $CRF_{k,j}(w,w')(p) = \frac{1}{T-p}\sum_{t=p+1}^{T} CRF_{k,j,t}(w,w')(p)$. When we additionally average over the potential outcomes, the average of the Horvitz and Thompson (1952) style estimator is an unbiased estimator of this object. We state this in the next proposition.

Proposition B.4. *Assume a potential outcome time series and that Assumption 4 holds. Then, if the moments exist,*

$$\mathbb{E}^{W,Y(\bullet)}[\hat{\bar{\tau}}_{k,j}(w,w')(p)] = C\bar{R}F_{k,j}(w,w')(p).$$

Moreover, if the moments are such that $\lim_{T\to\infty}\frac{1}{T-p}\sum_{t=p+1}^{T}\mathbb{E}^{Y(\bullet)}[\zeta^2_{k,j,t-p}(p)] < \infty$, *then* $\hat{\bar{\tau}}_{k,j}(w,w')(p) - CRF_{k,j}(w,w')(p) \xrightarrow{p} 0$.

Proof. The first result follows from an analogous argument as in Proposition B.2. Next, let $v_{k,j,t}(w,w')(p) = \hat{\bar{\tau}}_{k,j,t}(w,w')(p) - CRF_{k,j,t}(w,w')(p)$ and $v_{k,j}(w,w')(p) = \frac{1}{T-p}\sum_{t=p+1}^{T} v_{k,j,t-p}(w,w')(p)$. Then, as shown in Proposition B.2, $v_{k,j,t-p}$ is a martingale difference sequence. And so, note that, if the moments exist,

$$\bar{\zeta}^2_{k,j}(w,w')(p) = Var^{Y(\bullet)}\{\sqrt{T}v_{k,j}(w,w')(p)\} = \frac{T}{T-p}\frac{1}{T-p}\sum_{t=p+1}^{T}\mathbb{E}^{Y(\bullet)}[\zeta^2_{k,j,t-p}(p)].$$

Hence, so by the assumptions of the proposition, $\sqrt{T}v_{k,j}(w,w')(p) = O_p(1)$ over the non-anticipating treatment path and potential outcomes as $T \to \infty$. Since $v_{k,j}(w,w')(p) = \hat{\bar{\tau}}_{k,j}(w,w')(p) - CRF_{k,j}(w,w')(p)$, the second result follows. □ □

B.4 Additional results for Section 3.2.3: Nonparametric estimation with continuous treatments

Here we lay out some theory for the kernel estimator in the continuous treatment case. We will only consider the scalar treatment and outcome cases and so, we drop the subscripts k, j to simplify the notation.

Write the estimator as

$$\hat{\bar{\tau}}_t(w, w') = g_{t,p;h}(w) - g_{t,p;h}(w'),$$

where

$$g_{t,p;h}(w) = y_t(w^{obs}_{1:t-p-1}, W_{t-p:t}) \frac{\mathbb{1}(w^{obs}_{t-p} \in [w-h, w+h])}{F_{t-p}(w+h) - F_{t-p}(w'-h)}.$$

Define the compensator (which must exist, as we have conditioned on $Y_{1:T}(\bullet)$) and error

$$\mu_{t,p;h}(w) = \mathbb{E}\{g_{t,p;h}(w)|\mathcal{F}_{T,t-p-1}\}, \quad u_{t,p;h}(w) = g_{t,p;h}(w) - \mu_{t,p;h}(w).$$

as well as $\sigma^2_{t,p;h} := Var(\{u_{t,p;h}(w) - u_{t,p;h}(w')\}|\mathcal{F}_{T,t-p-1})$, which again must exist. By construction, $\mathbb{E}[u_{t,p;h}(w)|\mathcal{F}_{T,t-p-1}]$, so $u_{t,p;h}(w)$ is a martingale difference error with respect to $\mathcal{F}_{T,t-p-1}$. We introduce notation for the bias of the compensator as

$$b_{t,p;h}(w) = \mu_{t,p;h}(w) - \mathbb{E}\{y_t(w^{obs}_{1:t-p-1}, W_{t-p:t})|W_{t-p} = w, \mathcal{F}_{T,t-p-1}\}.$$

Then, $b_{t,p;h}(w) - b_{t,p;h}(w')$ is the conditional bias of $\hat{\bar{\tau}}_t(w, w')$ for $\bar{\tau}_t(w, w')$.

We now characterize this bias by separating out the role of h. To do this, $h \downarrow 0$ and so, the estimator is approaching the discrete case. To do this, we need the potential outcomes to be sufficiently smooth in order to define the following objects:

$$\dot{y}_t(w^{obs}_{1:t-p-1}, w, W_{t-p+1:t}) = \frac{\partial y_t(w^{obs}_{1:t-p-1}, w, W_{t-p+1:t})}{\partial w}, \quad \dot{\lambda}_t(w) = \mathbb{E}\{\dot{y}_t(w^{obs}_{1:t-p-1}, W_{t-p:t})|W_{t-p} = w, \mathcal{F}_{t-p-1}\}$$
$$\ddot{y}_t(w^{obs}_{1:t-p-1}, w, W_{t-p+1:t}) = \frac{\partial^2 y_t(w^{obs}_{1:t-p-1}, w, W_{t-p+1:t})}{\partial w^2}, \quad \ddot{\lambda}_t(w) = \mathbb{E}\{\ddot{y}_t(w^{obs}_{1:t-p-1}, W_{t-p:t})|W_{t-p} = w, \mathcal{F}_{t-p-1}\}$$

Denote conditional density of the treatment and its derivative as

$$f_{t-p}(w) = \frac{\partial F_{t-p}(w)}{\partial w}, \quad \dot{f}_{t-p}(w) = \frac{\partial^2 F_{t-p}(w)}{\partial w^2},$$

We assume that these exist. Then, as $h \downarrow 0$,

$$b_{t,p;h}(w) \simeq h^2 b_{t,p}(w), \quad \text{where} \quad b_{t,p}(w) = \frac{1}{3}\left\{\dot{\lambda}_t(w)\frac{\dot{f}_{t-p}(w)}{f_{t-p}(w)} + \frac{1}{2}\ddot{\lambda}_t(w)\right\}.$$

Why? First, from a second-order Taylor approximation, we have that that potential outcome is given by

$$y_t(w^{obs}_{1:t-p-1}, W_{t-p:t}) \approx \left\{y_t(w^{obs}_{1:t-p-1}, w, W_{t-p+1:t}) + (W_{t-p} - w)\,\dot{y}_t + \frac{1}{2}(W_{t-p} - w)^2\,\ddot{y}_t\right\}. \tag{24}$$

If $F_{t-p}(w)$ is twice continuously differentiable (as assumed), then $F_{t-p}(w+h) \approx F_{t-p}(w) + hf_{t-p}(w) + h^2\dot{f}(w)/2$, so $F_{t-p}(w+h) - F_{t-p}(w-h) = 2hf_{t-p}(w) + O(h^3)$. Then, we have that

$$\begin{aligned}\int_{-h}^{h} u\mathrm{d}F_{(W_t-w)|\mathcal{F}_{t-1}}(u) &= \int_{-h}^{h} uf_t(w+u)\mathrm{d}u \\ &\approx f_t(w)\int_{-h}^{h} u\mathrm{d}u + \dot{f}_t(w)\int_{-h}^{h} u^2\mathrm{d}u = \dot{f}_t(w)\frac{2h^3}{3},\end{aligned}$$

$$\int_{-h}^{h} u^2\mathrm{d}F_{(W_t-w)|\mathcal{F}_{t-1}}(u) = \int_{-h}^{h} u^2 f_t(w+u)\mathrm{d}u \simeq f_t(w)\int_{-h}^{h} u^2\mathrm{d}u = f_t(w)\frac{2h^3}{3}.$$

Now integrating the right hand side of Equation (24) times $\mathbb{1}(W^{obs}_{t-p} \in [w+h, w-h])$ with respect to $F_{W_{t-p}|\mathcal{F}_{t-p-1}}$ delivers the result.

The conditional variance is easier, it is $\sigma^2_{t,p;h} \approx \sigma^2_{t,p}h^{-1}$, where

$$\sigma^2_{t,p} = \frac{1}{2}\left\{\frac{\mathbb{E}\{y^2_t(w^{obs}_{1:t-p-1}, W_{t-p:t})|W_{t-p} = w, \mathcal{F}_{t-p-1}\}}{f_{t-p}(w)} + \frac{\mathbb{E}\{y^2_t(w^{obs}_{1:t-p-1}, W_{t-p:t})|W_{t-p} = w', \mathcal{F}_{t-p-1}\}}{f_{t-p}(w')}\right\}.$$

Why? From Equation (24), the first order term of the conditional variance of $y_t(w^{obs}_{1:t-p-1}, W_{t-p:t})$ will be

$$\begin{aligned}&\mathbb{E}\{y^2_t(w^{obs}_{1:t-p-1}, W_{t-p:t})|W_{t-p} = w, \mathcal{F}_{t-p-1}\}\tfrac{\{F_{t-p}(w+h)-F_{t-p}(w-h)\}[1-\{F_{t-p}(w+h)-F_{t-p}(w-h)\}]}{\{F_{t-p}(w+h)-F_{t-p}(w-h)\}^2} \\ &= \mathbb{E}\{y^2_t(w^{obs}_{1:t-p-1}, W_{t-p:t})|W_{t-p} = w, \mathcal{F}_{t-p-1}\}\tfrac{1-\{F_{t-p}(w+h)-F_{t-p}(w-h)\}}{F_{t-p}(w+h)-F_{t-p}(w-h)} \\ &\approx \mathbb{E}\{y^2_t(w^{obs}_{1:t-p-1}, W_{t-p:t})|W_{t-p} = w, \mathcal{F}_{t-p-1}\}\tfrac{1}{2hf_{t-p}(w)}.\end{aligned} \tag{25}$$

The analysis then delivers a traditional bias-variance tradeoff that chooses the bandwidth h to minimize the mean square error. If we have T observations, then the optimal bandwidth at lag p will scale as $h_p = c_p T^{-1/5}$, which implies the mean square error falls at rate $T^{-4/5}$.

B.5 Additional results for section 5: Local projections

In this subsection, we collect a series of additional results that further contextualize local projections and local projections with an instrument in the context of a shocked potential outcome time series.

B.5.1 Section 5.1: Local projection with observed treatments

The following result shows that the ordinary least squares estimator identifies a time-average of the k, j-th lag-p time-t causal coefficients if the causal effects are additive but not I-additive.

Proposition B.5. *Maintain the assumptions provided in Theorem 5.1. Additionally assume that*

D. $W_{j,t-p}$ *is a stationary stochastic process and that the auto-covariances of* $W^2_{j,t-p}$ *are absolutely summable. That is,* $\lim_{n\to\infty}\sum_{m=-n}^{n}|\gamma_{j,m}| < \infty$ *with* $Cov(W^2_{j,t-p}, W^2_{j,t-p+m}) = \gamma_{j,m}$.

E. The causal coefficients are absolutely summable with $\sum_{t=p+1}^{\infty}|\beta_{p,k,j,t}| < \infty$.

Define $\beta^*_{p,k,j} := \lim_{T\to\infty}\frac{1}{T-p-1}\sum_{t=p+1}^{T}\beta_{p,k,j,t}$. *Then,* $\hat{\beta}^{OLS}_{p,k,j} \xrightarrow{p} \beta^*_{p,k,j}$.

Proof. This proof is adapted from an argument given in Section 12.4 of Theil (1971). Under the assumptions of Theorem 5.1, the causal regression of $Y_{k,t}$ on $W_{j,t-p}$ is

$$Y_{k,t} = \alpha + \beta_{p,k,j,t}W_{j,t-p} + \eta_{k,j,t}.$$

Adding and subtracting $\bar{\beta}_{p,k,j,T}W_{j,t-p}$, rewrite this as

$$Y_{k,t} = \alpha + \bar{\beta}_{p,k,j,T}W_{j,t-p} + \nu_{k,j,t} \text{ where,}$$

$$\nu_{k,j,t} = (\beta_{p,k,j,t} - \bar{\beta}_{p,k,j,T})W_{j,t-p} + \eta_{k,j,t}.$$

The OLS estimator is then

$$\begin{aligned}
\hat{\beta}^{OLS}_{p,k,j} &= \frac{\sum_{t=p+1}^{T} Y_{k,t}(W_{j,t-p} - \bar{W}_j)}{\sum_{t=p+1}^{T}(W_{j,t-p} - \bar{W}_j)^2} \\
&= \bar{\beta}_{p,k,j,T} + \frac{\frac{1}{T-p-1}\sum_{t=p+1}^{T}(\beta_{p,k,j,t} - \bar{\beta}_{p,k,j,T})W_{j,t-p}(W_{j,t-p} - \bar{W}_j)}{\frac{1}{T-p-1}\sum_{t=p+1}^{T}(W_{j,t-p} - \bar{W}_j)^2} + \frac{\frac{1}{T-p-1}\sum_{t=p+1}^{T}\eta_{k,j,t}(W_{j,t-p} - \bar{W}_j)}{\frac{1}{T-p-1}\sum_{t=p+1}^{T}(W_{j,t-p} - \bar{W}_j)^2}.
\end{aligned}$$

As in Theorem 5.1, the third term converges to zero as $\eta_{k,j,t}$ is uncorrelated with $W_{j,t-p}$. Under the additional assumptions, the second term converges to zero as well. Rewrite the numerator of the seceond term as

$$\frac{1}{T-p-1}\sum_{t=p+1}^{T}\beta_{p,k,j,t}W_{j,t-p}(W_{j,t-p}-\bar{W}_j)-\bar{\beta}_{p,k,j,T}\frac{1}{T-p-1}\sum_{t=p+1}^{T}W_{j,t-p}(W_{j,t-p}-\bar{W}_j)$$
$$=\frac{1}{T-p-1}\sum_{t=p+1}^{T}\beta_{p,k,j,t}W_{j,t-p}^2-\bar{W}_j\frac{1}{T-p-1}\sum_{t=p+1}^{T}\beta_{p,k,j,t}W_{j,t-p}$$
$$-\bar{\beta}_{p,k,j,T}\frac{1}{T-p-1}\sum_{t=p+1}^{T}W_{j,t-p}^2+\bar{\beta}_{p,k,j,T}\bar{W}_j\frac{1}{T-p-1}\sum_{t=p+1}^{T}W_{j,t-p}$$

Since $W_{j,t-p}$ is a martingale difference sequence, it is serially uncorrelated. Therefore, using Assumption D, a LLN can be used to show

$$\frac{1}{T-p-1}\sum_{t=p+1}^{T}W_{j,t-p}-\frac{1}{T-p-1}\sum_{t=p+1}^{T}\mathbb{E}[W_{j,t-p}]=\frac{1}{T-p-1}\sum_{t=p+1}^{T}W_{j,t-p}\xrightarrow{p}0,$$
$$\frac{1}{T-p-1}\sum_{t=p+1}^{T}\beta_{p,k,j,t}W_{j,t-p}-\frac{1}{T-p-1}\sum_{t=p+1}^{T}\beta_{p,k,j,t}\mathbb{E}[W_{j,t-p}]=\frac{1}{T-p-1}\sum_{t=p+1}^{T}\beta_{p,k,j,t}W_{j,t-p}\xrightarrow{p}0.$$

Next, consider the third term. Using Assumption D, we can also apply a LLN and so,

$$\frac{1}{T-p-1}\sum_{t=p+1}^{T}W_{j,t-p}^2\xrightarrow{p}\sigma_j^2$$
$$\bar{\beta}_{p,k,j,T}\frac{1}{T-p-1}\sum_{t=p+1}^{T}W_{j,t-p}^2\xrightarrow{p}\beta^*_{p,k,j}\sigma_j^2.$$

Finally, consider the first term. Take its expectation and obtain that

$$\mathbb{E}^W[\frac{1}{T-p-1}\sum_{t=p+1}^{T}\beta_{p,k,j,t}W_{j,t-p}^2]=\frac{1}{T-p-1}\sum_{t=p+1}^{T}\beta_{p,k,j,t}\mathbb{E}^W[W_{j,t-p}^2]=\frac{1}{T-p-1}\sum_{t=p+1}^{T}\beta_{p,k,j,t}\sigma_j^2$$

Now consider the variance

$$\mathbb{E}^W[\{\frac{1}{T-p-1}\sum_{t=p+1}^{T}\beta_{p,k,j,t}W_{j,t-p}^2-\frac{1}{T-p-1}\sum_{t=p+1}^{T}\beta_{p,k,j,t}\sigma_j^2\}^2]$$

$$
\begin{aligned}
&= \mathbb{E}^W[\{\frac{1}{T-p-1}\sum_{t=p+1}^{T}\beta_{p,k,j,t}(W_{j,t-p}^2-\sigma_j^2)\}^2] \\
&= \frac{1}{(T-p-1)^2}\sum_{t=p+1}^{T}\sum_{s=p+1}^{T}\beta_{p,k,j,t}\beta_{p,k,j,s}\mathbb{E}^W[(W_{j,t-p}^2-\sigma_j^2)(W_{j,s-p}^2-\sigma_j^2)] \\
&= \frac{1}{(T-p-1)^2}\sum_{t=p+1}^{T}\sum_{s=p+1}^{T}\beta_{p,k,j,t}\beta_{p,k,j,s}\gamma_{j,t-s}.
\end{aligned}
$$

Under Assumptions D, E, this converges to zero. So, with this, apply a LLN to obtain

$$
\frac{1}{T-p-1}\sum_{t=p+1}^{T}\beta_{p,k,j,t}W_{j,t-p}^2 - \frac{1}{T-p-1}\sum_{t=p+1}^{T}\beta_{p,k,j,t}\sigma_j^2 \xrightarrow{p} 0.
$$

It immediately follows that $\frac{1}{T-p-1}\sum_{t=p+1}^{T}\beta_{p,k,j,t}W_{j,t-p}^2 \xrightarrow{p} \beta_{p,k,j}^*\sigma_j^2$. So, we conclude that the second term in the OLS estimator written above converges in probability to zero and the result follows. □

Next, we ask the question: Can local projections be interpreted as providing an additive approximation to general time-t causal effects? Here we establish one such approximation based on the best approximation to the causal effects in an L^2 sense. Following the approach taken in Section 3, we fix the treatments to follow the observed path $w_{1:t}^{obs}$ until period $t-p-1$. In the next proposition, we fix the potential outcomes and thereby provide a linear approximation to the possibly non-linear weighted causal effect in Definition 7. Recall that $\mathcal{F}_{T,t}$ denotes the history $\{w_{1:t}^{obs}, Y_{1:T}(\bullet)\}$.

Proposition B.6 (Projection of causal effects)**.** *Assume a shocked potential outcome time series. Let $W_{1:t} = (w_{1:t-p-1}^{obs}, W_{t-p:t})$ and fix the counter-factual path as $w'_{1:t} = (w_{1:t-p-1}^{obs}, \bar{w}_{t-p:t})$. Denote*

$$
\sigma_{j,T,t-p}^2 := \mathbb{E}^{W\,|\,Y(\bullet)}[W_{j,t-p}^2 \,|\, \mathcal{F}_{T,t-p-1}], \tag{26}
$$

$$
d_{k,j,T,t}(p) := \mathbb{E}^{W\,|\,Y(\bullet)}[Y_{k,t}(w_{1:t-p-1}, W_{t-p:t})W_{j,t-p} \,|\, \mathcal{F}_{T,t-p-1}], \tag{27}
$$

$$
\beta_{k,j,T,t}^*(p) := d_{k,j,T,t}(p)/\sigma_{j,T,t-p}^2, \tag{28}
$$

$$
\alpha_{k,T,t}^*(p) := \mathbb{E}^{W\,|\,Y(\bullet)}[\tau_{k,t}(W_{1:t}, w'_{1:t}) \,|\, \mathcal{F}_{T,t-p-1}]. \tag{29}
$$

Assume that $\sigma_{j,T,t-p}^2 > 0$. Then,

$$
\{\alpha_{k,T,t}^*(p), \beta_{k,j,T,t}^*(p)\} = \arg\min_{a,b}\mathbb{E}^{W\,|\,Y(\bullet)}\left[\{\tau_t(W_{1:t}, w'_{1:t}) - a - bW_{j,t-p}\}^2 \,|\, \mathcal{F}_{T,t-p-1}\right], \tag{30}
$$

$$
\beta_{k,j,T,t}^*(p) = \arg\min_{b}\mathbb{E}^{W\,|\,Y(\bullet)}\left[\{Y_t(w_{1:t-p-1}^{obs}, W_{t-p:t}) - bW_{j,t-p}\}^2 \,|\, \mathcal{F}_{T,t-p-1}\right]. \tag{31}
$$

Proof. Begin by noting that

$$\mathbb{E}^{W|Y(\bullet)}[\tau_{k,t}(W_{1:t}, W'_{1:t})|\mathcal{F}_{T,t-p-1}] = \mathbb{E}^{W|Y(\bullet)}[Y_{k,t}(w^{obs}_{1:t-p-1}, W_{t-p:t})|\mathcal{F}_{T,t-p-1}] - Y_{k,t}(w^{obs}_{1:t-p-1}, \bar{w}_{t-p:t}).$$

Now let $c_{k,t,p} := Y_{k,t}(w^{obs}_{1:t-p-1}, \overline{w}_{t-p:t})$ and

$$S_{k,j,T,t,p}(a,b) := \mathrm{E}^{W|Y(\bullet)}\left[\left\{Y_{k,t}(w^{obs}_{1:t-p-1}, W_{t-p:t}) - Y_{k,t}(w^{obs}_{1:t-p-1}, \overline{w}_{t-p:t}) - a - bW_{j,t-p}\right\}^2 |\mathcal{F}_{T,t-p-1}\right] \tag{32}$$

Given $\mathcal{F}_{T,t-p-1}$, $c_{k,t,p}$ it is non-stochastic. Then,

$$\begin{aligned}
S_{k,j,T,t,p}(a,b) &= \mathrm{E}^{W|Y(\bullet)}\left[\left\{Y_{k,t}(w^{obs}_{1:t-p-1}, W_{t-p:t}) - \{c_{k,t,p} + a\} - bW_{j,t-p}\right\}^2 |\mathcal{F}_{T,t-p-1}\right] \\
&= \mathrm{E}^{W|Y(\bullet)}\left[\left\{Y_{k,t}(w^{obs}_{1:t-p-1}, W_{t-p:t}) - bW_{j,t-p}\right\}^2 |\mathcal{F}_{T,t-p-1}\right] \\
&- 2\{c_{k,t,p} + a\}\, \mathrm{E}^{W|Y(\bullet)}\left[Y_{k,t}(w^{obs}_{1:t-p-1}, W_{t-p:t}) - bW_{j,t-p}|\mathcal{F}_{T,t-p-1}\right] + \{c_{k,t,p} + a\}^2 \\
&= \mathrm{E}^{W|Y(\bullet)}\left[\left\{Y_{k,t}(w^{obs}_{1:t-p-1}, W_{t-p:t}) - bW_{j,t-p}\right\}^2 |\mathcal{F}_{T,t-p-1}\right] \\
&- 2\{c_{k,t,p} + a\}\, \mathrm{E}^{W|Y(\bullet)}\left[Y_{k,t}(w^{obs}_{1:t-p-1}, W_{t-p:t})|\mathcal{F}_{T,t-p-1}\right] + \{c_{k,t,p} + a\}^2,
\end{aligned}$$

where $\mathrm{E}^{W|Y(\bullet)}[W_{j,t-p}|\mathcal{F}_{T,t-p-1}] = 0$ as the treatment is a shock. So,

$$\begin{aligned}
\beta^*_{k,j,T,t}(p) &= \arg\min_b S_{k,j,T,t,p}(a,b) \\
&= \arg\min_b \mathrm{E}^{W|Y(\bullet)}\left[\left\{Y_{k,t}(w^{obs}_{1:t-p-1}, W_{t-p:t}) - bW_{j,t-p}\right\}^2 |\mathcal{F}_{T,t-p-1}\right] \\
&= d_{k,j,T,t}(p)/\sigma^2_{j,T,t-p}
\end{aligned}$$

by a standard derivation of the linear projection coefficient (e.g. see the textbook exposition on pg. 33 of Hansen (2018)). Moreover,

$$\begin{aligned}
\alpha_{k,T,t}(p) &= \arg\min_a S_{T,t,p}(a,b) \\
&= \mathrm{E}^{W|Y(\bullet)}\left[Y_{k,t}(w^{obs}_{1:t-p-1}, W_{t-p:t})|\mathcal{F}_{T,t-p-1}\right] - c_{k,t,p}.
\end{aligned}$$

□

As shorthand, we write the approximation[20] defined in Proposition B.6 as

$$Proj(Y_t(w^{obs}_{1:t-p-1}, W_{t-p:t}) \,|\, 1, W_{j,t-p}, \mathcal{F}_{T,t-p-1}) = \alpha^*_{k,T,t}(p) + \beta^*_{k,j,T,t}(p) W_{j,t-p}. \tag{33}$$

In the next proposition, we drop the conditioning on $Y_{1:t}(\bullet)$. The approximation now averages over both the treatment paths and the potential outcomes. As a result, it now provides a linear approximation to the causal response function in Definition 8. Again recall that $\mathcal{F}_t$ denotes the history $\{w^{obs}_{1:t}, Y_{1:t}(\bullet)\}$.

Proposition B.7. *Assume a shocked potential outcome time series. Let* $W_{1:t} = (w^{obs}_{1:t-p-1}, W_{t-p:t})$ *and fix the counter-factual path as* $w'_{1:t} = (w^{obs}_{1:t-p-1}, \bar{w}_{t-p:t})$. *Assume that*

$$\sigma^2_{j,t-p} := \mathbb{E}^{W,Y(\bullet)}[W^2_{j,t-p} \,|\, \mathcal{F}_{t-p-1}], \tag{34}$$

$$d_{k,j,t}(p) := \mathbb{E}^{W,Y(\bullet)}[Y_{k,t}(w_{1:t-p-1}, W_{t-p:t}) W_{j,t-p} \,|\, \mathcal{F}_{t-p-1}], \tag{35}$$

$$\alpha^*_{k,t}(p) := \mathbb{E}^{W,Y(\bullet)}[\tau_{k,t}(W_{1:t}, w'_{1:t}) \,|\, \mathcal{F}_{t-p-1}] \tag{36}$$

exist and that $\sigma^2_{j,t-p} > 0$. *Define:* $\beta^*_{k,j,t}(p) := d_{k,j,T,t}(p)/\sigma^2_{j,t-p}$. *Then, assuming the moments exist,*

$$\{\alpha^*_{k,t}(p), \beta^*_{k,j,t}(p)\} = \arg\min_{a,b} \mathbb{E}^{W,Y(\bullet)}\left[\{\tau_t(W_{1:t}, w'_{1:t}) - a - bW_{j,t-p}\}^2 \,|\, \mathcal{F}_{t-p-1}\right] \tag{37}$$

$$\beta^*_{k,j,t}(p) = \arg\min_{b} \mathbb{E}^{W,Y(\bullet)}\left[\{Y_t(w^{obs}_{1:t-p-1}, W_{t-p:t}) - bW_{j,t-p}\}^2 \,|\, \mathcal{F}_{t-p-1}\right]. \tag{38}$$

Proof. Define

$$S_{k,j,t,p}(a,b) := \mathrm{E}^{W,Y(\bullet)}\left[\left\{Y_{k,t}(w^{obs}_{1:t-p-1}, W_{t-p:t}) - Y_{k,t}(w^{obs}_{1:t-p-1}, \overline{w}_{t-p:t}) - a - bW_{j,t-p}\right\}^2 |\mathcal{F}_{t-p-1}\right] \tag{39}$$

and

$$c_{k,t,p} := \mathrm{E}^{W,Y(\bullet)}\left\{Y_{k,t}(w^{obs}_{1:t-p-1}, \overline{w}_{t-p:t}) |\mathcal{F}_{t-p-1}\right\}. \tag{40}$$

Given $\mathcal{F}_{t-p-1}$, $c_{k,t,p}$ is non-stochastic. Then,

$$S_{k,j,t,p}(a,b) = \mathrm{E}^{W,Y(\bullet)}\left[\left\{Y_{k,t}(w^{obs}_{1:t-p-1}, W_{t-p:t}) - \{c_{k,t,p} + a\} - bW_{j,t-p}\right\}^2 |\mathcal{F}_{t-p-1}\right]$$

[20]This approximation is defined with respect to the $W_{j,t-p}$. This connects to common empirical practice. It is straightforward to extend this result to a vector of treatments.

$$
\begin{aligned}
&= \mathrm{E}^{W,Y(\bullet)}\left[\left\{Y_{k,t}(w^{obs}_{1:t-p-1}, W_{t-p:t}) - bW_{j,t-p}\right\}^2 |\mathcal{F}_{t-p-1}\right] \\
&- 2\left\{c_{k,t,p} + a\right\} \mathrm{E}^{W,Y(\bullet)}\left[Y_{k,t}(w^{obs}_{1:t-p-1}, W_{t-p:t}) - bW_{j,t-p}|\mathcal{F}_{t-p-1}\right] + \left\{c_{k,t,p} + a\right\}^2 \\
&= \mathrm{E}^{W,Y(\bullet)}\left[\left\{Y_{k,t}(w^{obs}_{1:t-p-1}, W_{t-p:t}) - bW_{j,t-p}\right\}^2 |\mathcal{F}_{t-p-1}\right] \\
&- 2\left\{c_{k,t,p} + a\right\} \mathrm{E}^{W,Y(\bullet)}\left[Y_{k,t}(w^{obs}_{1:t-p-1}, W_{t-p:t})|\mathcal{F}_{t-p-1}\right] + \left\{c_{k,t,p} + a\right\}^2,
\end{aligned}
$$

where $\mathrm{E}^{W,Y(\bullet)}(W_{j,t-p}|\mathcal{F}_{t-p-1}) = 0$ as the treatment is a shock. So, as in the Proof of Proposition B.6, we have

$$
\begin{aligned}
\beta^*_{k,j,t}(p) &= \arg\min_b S_{k,j,t,p}(a,b) \\
&= \arg\min_b \mathrm{E}^{W,Y(\bullet)}\left[\left\{Y_t(w^{obs}_{1:t-p-1}, W_{t-p:t}) - bW_{t-p}\right\}^2 |\mathcal{F}_{t-p-1}\right] \\
&= d_{k,j,t}(p)/\sigma^2_{j,t-p}
\end{aligned}
$$

and moreover,

$$
\begin{aligned}
\alpha^*_{k,t}(p) &= \arg\min_a S_{t,p}(a,b) \\
&= \mathrm{E}^{W,Y(\bullet)}\left[Y_t(w^{obs}_{1:t-p-1}, W_{t-p:t})|\mathcal{F}_{t-p-1}\right] - c_{k,t,p}.
\end{aligned}
$$

□

Once again, we write this in shorthand as

$$
Proj(Y_{k,t}(w^{obs}_{1:t-p-1}, W_{t-p:t}) \,|\, 1, W_{j,t-p}, \mathcal{F}_{t-p-1}) = \alpha^*_{k,t}(p) + \beta^*_{k,j,t}(p)W_{j,t-p}. \tag{41}
$$

If the shocked potential outcome time series is covariance stationary, then the time subscripts on all of these objects can be dropped, as can be the $\mathcal{F}_{t-p-1}$. Then we can similarly obtain a linear approximation to the impulse response function following the construction in Proposition B.6 and Proposition B.7. The result is

$$
Proj(Y_{k,t}|1, W_{j,t-p}) = \alpha^*_k(p) + \beta^*_{k,j}(p)W_{j,t-p}, \tag{42}
$$

where

$$\alpha_k^*(p) = \mathbb{E}^{W,Y(\bullet)}[\tau_{k,t}], \quad \beta_{k,j}^*(p) = \mathbb{E}^{W,Y(\bullet)}[Y_{k,t}W_{j,t-p}]/\mathbb{E}^{W,Y(\bullet)}[W_{j,t-p}^2]. \tag{43}$$

B.5.2 Section 5.2: Local projection with an instrumental variable

In the existing LP-IV literature, it is typical to consider a weaker assumption than Assumption B in Theorem 5.2. Instead of assuming that the treatment and proxy are a joint martingale difference sequences, it is common to assume that the proxy satisfies *lead-lag exogeneity.* However, this assumption alone is not sufficient to identify a dynamic causal effect in the potential outcome time series framework. We formalize this in the following proposition

Proposition B.8 (LP-IV with lead-lag exogeneity)**.** *Assume a proxy potential outcome time series as in Definition 13 and additionally that*

A. The causal effects are additive, given in Definition 6.

B Treatment and proxy satisfy:

1. *Treatment is a shock, given in Definition 10.*
2. $\mathbb{E}^{W,\hat{W}}[\hat{W}_s W_t] = 0$ *for* $s \neq t$*.*
3. $\mathbb{E}^{Y,\hat{W}}[Y_t(\bar{w}_{1:t})\hat{W}_{t-p}] = 0$*, for* $p \geq 0$*.*

C. $\mathbb{E}^{W,\hat{W}}[W_t\hat{W}_s]$ *and* $\mathbb{E}^{Y,\hat{W}}[Y_t(w_{1:t})\hat{W}_{t-p}]$ *exists for known non-stochastic counterfactual path.*

D. $\mathbb{E}^{W,\hat{W}}[W_{i,t}\hat{W}_{j,t}] = 1_{i=j}\gamma_j$*.*

Then, for $p \geq 0$*,*

$$\mathbb{E}^{Y,W,\hat{W}}[Y_{k,t}(W_{1:t})\hat{W}_{j,t-p}] = \beta_{p,k,j,t}\gamma_j. \tag{44}$$

Proof. The proof proceeds as before in Theorem 5.2. □

From Proposition B.8, we make the following remarks.

1. Assumption B2 is typically referred to as "lead-lag exogeneity" in the existing literature. It is not sufficient to identify the causal coefficient in the potential outcome time series. Without further assumptions, we are left with

   $$\mathbb{E}^{Y,W,\hat{W}}[Y_{k,t}(W_{1:t})\hat{W}_{j,t-p}] = \beta_{p,k,j,t}\gamma_j + \mathbb{E}^{Y,\hat{W}}[Y_t(\bar{w}_{1:t})\hat{W}_{t-p}], \tag{45}$$

   which is not causal.

2. Assumption B3 is an additional, necessary restriction that is required to causally interpret the LP-IV estimand. Assumption B in Theorem 5.2 implies that Assumption B2 and Assumption B3 in Theorem **??** hold. This is stated in Lemma 5.1.

3. Assumption B3 is a restriction on the joint distribution of the unobserved counterfactual and the proxy. As a result, it is a difficult assumption to verify in applications. Consider the monetary policy shock example. If $\hat{W}_t$ is a proxy for the monetary policy shock, Assumption B3 requires that the proxy be uncorrelated with the counterfactual unemployment rate at time t.

Proposition B.8 illustrates the value of the potential outcome time series framework as we have used it to identify a condition that is missing in the literature and is necessary for the causal validity of LP-IV under existing assumptions.

We now ask – what does the IV estimator identify if the causal effects are additive but not I-additive?

Proposition B.9. *Maintain the assumptions provided in Theorem 5.2. Additionally assume that*

E. $(W_{j,t-p}, \hat{W}_{j,t-p})$ *are jointly stationary.*

F. *The causal coefficients are absolutely summable with* $\sum_{t=p+1}^{\infty} |\beta_{p,k,j,t}| < \infty$.

Then, the IV estimator identifies $\beta^*_{p,k,j}$ *as* $T \to \infty$*. That is,*

$$\frac{\sum_{t=p+1}^{T} Y_{k,t}\hat{W}_{j,t-p}}{\sum_{t=p+1}^{T} W_{j,t-p}\hat{W}_{j,t-p}} \xrightarrow{p} \beta^*_{p,k,j}.$$

Proof. The argument is analogous to Proposition B.5. Under the assumptions of Theorem 5.2, the causal regression of $Y_{k,t}$ on $W_{j,t-p}$ is

$$Y_{k,t} = \beta_{p,k,j,t} W_{j,t-p} + \eta_{k,j,t}.$$

Add and subtract $\bar{\beta}_{p,k,j} W_{j,t-p}$ to write

$$Y_{k,t} = \bar{\beta}_{p,k,j} W_{j,t-p} + \{(\beta_{p,k,j,t} - \bar{\beta}_{p,k,j}) W_{j,t-p} + \eta_{k,j,t}\}.$$

Consider the numerator of the IV estimator

$$\frac{1}{T-p-1}\sum_{t=p+1}^{T} Y_{k,t}\hat{W}_{j,t-p} = \bar{\beta}_{p,k,j}\{\frac{1}{T-p-1}\sum_{t=p+1}^{T} W_{j,t-p}\hat{W}_{j,t-p}\}$$

$$+\frac{1}{T-p-1}\sum_{t=p+1}^{T}(\beta_{p,k,j,t}-\bar{\beta}_{p,k,j})W_{j,t-p}\hat{W}_{j,t-p}$$
$$+\frac{1}{T-p-1}\sum_{t=p+1}^{T}\eta_{k,j,t}\hat{W}_{j,t-p}.$$

Under Assumption E, apply a LLN to the first and third terms to get

$$\bar{\beta}_{p,k,j}\{\frac{1}{T-p-1}\sum_{t=p+1}^{T}W_{j,t-p}\hat{W}_{j,t-p}\}\xrightarrow{\beta}{}^{*}_{p,k,j}\,\gamma_j$$
$$\frac{1}{T-p-1}\sum_{t=p+1}^{T}\eta_{k,j,t}\hat{W}_{j,t-p}\xrightarrow{p}0.$$

Consider the second term. Assumption F is sufficient to conclude that

$$\frac{1}{T-p-1}\sum_{t=p+1}^{T}\beta_{p,k,j,t}W_{j,t-p}\hat{W}_{j,t-p}\xrightarrow{p}\beta^{*}_{p,k,j}\gamma_j$$
$$\frac{1}{T-p-1}\sum_{t=p+1}^{T}\bar{\beta}_{p,k,j,t}W_{j,t-p}\hat{W}_{j,t-p}\xrightarrow{p}\beta^{*}_{p,k,j}\gamma_j.$$

As a result, the second term also converges in probability to zero. The result follows. □

Notice that in this result, we still restricted that the correlation between the proxy $\hat{W}_{j,t}$ and the treatment $W_{j,t}$ to be time invariant – that is, $\mathbb{E}^{W,\hat{W}}[W_{j,t}\hat{W}_{j,t}]=\gamma_j$ does not depend on t. This may be a reasonable assumption for empirical applications of the LP-IV method. For example, consider the example in Stock and Watson (2018) in which high frequency changes in the rates on Federal Funds futures contracts are used to proxy for the monetary policy shock around policy announcements. It may be plausible to assume that the quality of this proxy does not vary over time but the effects of monetary policy on real outcomes (such as unemployment) varies.

B.6 Monte Carlo simulations

In this section, we conduct a simple Monte Carlo simulation to assess the performance of the non-parametric estimator of the weighted causal effect that is used in the empirical application in Section 6.

The Monte Carlo exercise is as follows. Let $\hat{\beta}_{k,j,p}$ denote the estimated coefficients from the local projection estimator described in Section 6.1 and let $\tilde{W}_{j,t-p}\sim N(\hat{\gamma},\hat{\sigma}^2)$, where $\hat{\gamma},\hat{\sigma}^2$ are the parameters governing the distribution of the observed treatments that are estimated via maximum likelihood. Note that treatments are assumed to be independent across time periods. We simulate outcomes from the following moving average model

$$Y_{k,t}^{b} = \hat{\beta}_{k,j,0}\tilde{W}_{j,t}^{b} + \hat{\beta}_{k,j,1}\tilde{W}_{j,t-1}^{b} + \ldots + \hat{\beta}_{k,j,P}\tilde{W}_{j,t-P}^{b} \quad \text{for } b = 1, \ldots B, \tag{46}$$

where P denotes the maximum horizon of the causal effects and B is the number of simulations. To match the local projection estimator, we set $P = 24$. Note that the simulated outcomes have I-additive causal effects with causal coefficients given by the estimated local projection coefficients $\hat{\beta}_{k,j,p}$. As a result, the weighted causal effects are known and simply equal the causal coefficients.

To assess its performance of the non-parametric weighted causal effects estimator, we estimate the non-parametric weighted causal effects estimator described in Section 6.2 on the simulated data for treatment $W = 0.1$ and counterfactual $W' = -0.1$. In particular, we construct the bandwidth using the heuristic discussed in Section 3.2.3, which approximates the bias and variance of the estimator under the assumption of additive causal effects. If the estimator performs well, we would therefore expect that the estimated weighted causal effect $\hat{\bar{\tau}}_{k,j}(w, w')(p)$ would be roughly equal to $\hat{\beta}_{k,j,p}/5$. To check this, we conduct the Monte Carlo simulation for $B = 10,000$ and for each outcome series used in Section 6. The results are summarized in Figure 3 below, in which we plot the average estimate of the weighted causal effect across all simulations as well as $\hat{\beta}_{k,j,p}/5$. As can be seen, the weighted causal effect estimator is approximately unbiased at each horizon. We additionally plot the 2.5% and 97.5% quantiles of the estimated weighted causal effects.

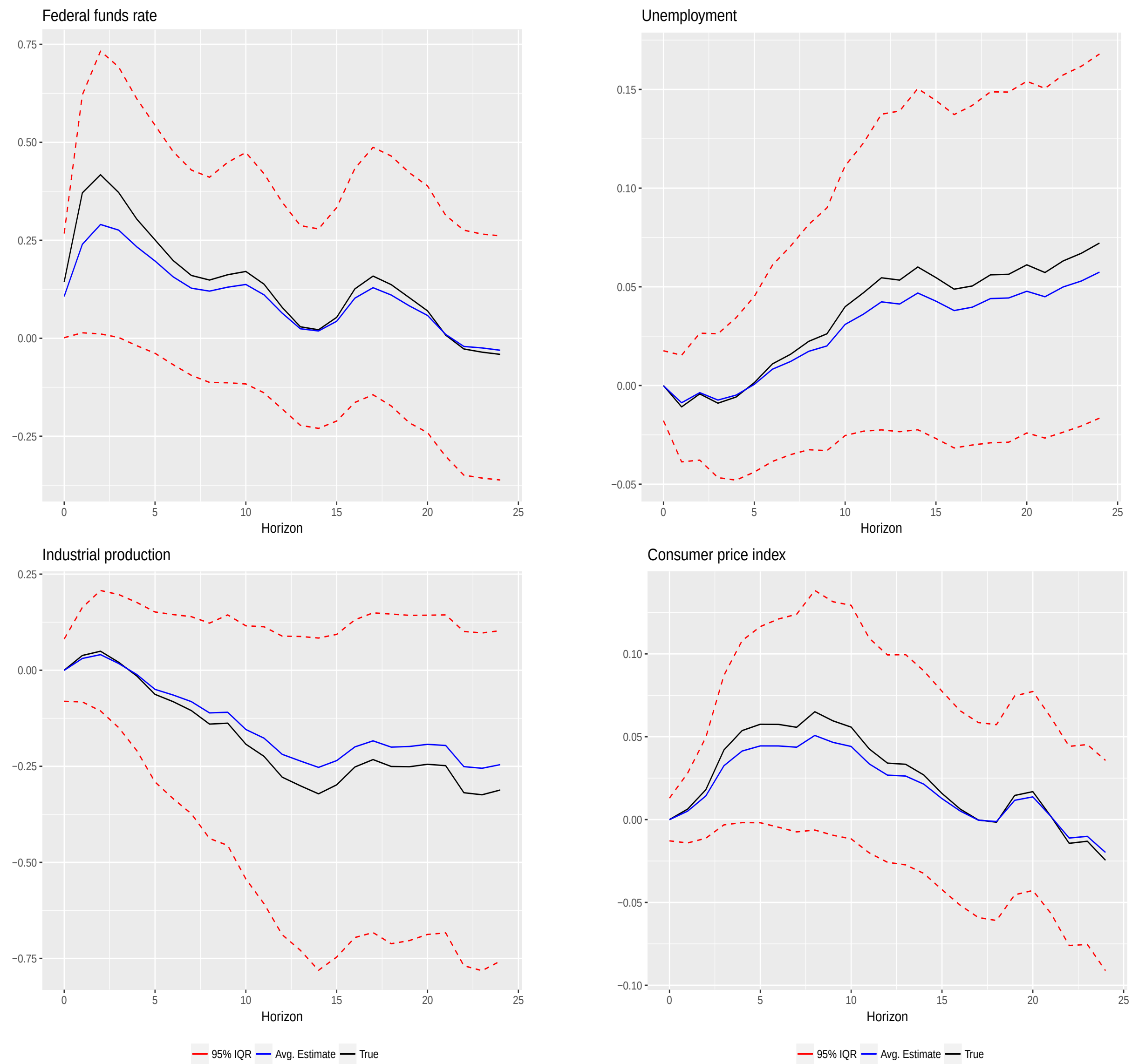

Figure 3: Monte Carlo simulation results. *Notes:* The solid black line plots the true causal coefficients divided by 5 that are used to simulate the data. The solid blue line plots the average estimated weighted causal effect $\bar{\tau}_{k,j}(w, w')(p)$ for $p = 1, \ldots, 24$ across 10,000 Monte Carlo simulations. The dotted red lines plot the 2.5% quantile and the 97.5% quantiles of the estimated weighted causal effects.