Uncovering Nonlinearities with Regression Anatomy: Online Appendix

Paul Bousquet

February 14, 2025

Full Expansion of FWL

Recall the general form from Section 2 of the paper

$$\omega_i = \frac{\text{cov}(1_{a \leq \varepsilon_t}, X_i^{\perp})}{\text{var}(X_i^{\perp})}$$

where $X_i^{\perp}$ is the residual from regressing X_i on the other elements in $\boldsymbol{X}_t$. We can unpack this definition to get things soley in terms of covariances and variance of terms of $\boldsymbol{X}_t$, which amounts to an expansion of the FWL theorem. To my knowledge, this expansion has not been done previously and for good reason – the full form amounts to several messy recursions that offer absolutely no insight to write out. However to motivate the use of deep learning to address one of the central issues in this paper, it may be useful to see why it's difficult to conjure up functional forms that will produce appropriate weighting.

For what follows, consider $\boldsymbol{X}$ to be a generic matrix of N covariates in a regression (which can include a vector of 1s) and X_i to be its i-th element. Keeping with notation from earlier, $X_i^{\perp}$ is the residual from X_i on the remaining elements of $\boldsymbol{X}$. WLOG, we will initially look at an example where $i = 1$. Further consider $X_n^{\perp_1}$ to be regressing the n-th element of $\boldsymbol{X}$ on its the remaining parts <u>excluding X_1</u>. Then

$$X_1^{\perp} = X_1 - \sum_{n=2}^{N} \frac{\text{cov}(X_1, X_n^{\perp_1})}{\text{var}(X_n^{\perp_1})} X_n$$

We can keep unpacking these terms but it should be clear that indexing is quickly going to become a nightmare because the "exclusions" will not be in a consistent ordering across the components (and sub-components, and sub-sub-components,...) of this summation. Things would have already got a bit messy notation wise had we done a formula for a generic $X_i^{\perp}$. So we will have to break this up into several parts.

First, we will deal with the covariance terms and keep the variance terms fixed. Again using $i = 1$ as the most coherent example and only focusing on the first term term in the sum above ($n = 2$), if we unpack a bit more we will have

$$\frac{\text{cov}(X_1, X_n^{\perp_1})}{\text{var}(X_n^{\perp_1})} = \frac{\text{cov}(X_1, X_2) - \frac{\text{cov}(X_2, X_3^{\perp_{1:2}})}{\text{var}(X_3^{\perp_{1:2}})} \text{cov}(X_1, X_3) + \ldots}{\text{var}(X_2^{\perp_1})}$$

where $X_3^{\perp_{1:2}}$ are the residuals of regressing X_3 on the remaining elements of $\boldsymbol{X}$ excluding X_1, X_2. To even coherently define the remaining terms in the numerator, new notation has to be introduced to deal with the order in which variables are excluded from the "sub-regressions". To address this, we will define things in chunks. Again keeping with the $i = 1$ and $n = 2$ case because it's the cleanest, note that

$$\frac{\text{cov}(X_1, X_2^{\perp_1})}{\text{var}(X_2^{\perp_1})} = \frac{C_{1,2} - C_{1,3}\left(\frac{C_{2,3}}{V_{2,3}} - C_{2,4}\left(\frac{C_{3,4}}{V_{3,4}V_{2,3}} - C_{3,5}(\ldots) - \ldots\right) - \ldots\right) + \ldots}{V_{1,2}}$$

where $C_{p,q}$ represents the covariance between elements p and q of $\boldsymbol{X}$ and $V_{p,q}$ is $\text{var}(X_q^{\perp_{1:p}})$. Ignoring the unavoidably ugly denoting of what terms correspond to, we can see a light of coherency at the end of the recursion tunnel. While there may now be a decipherable pattern to latch onto, this hasn't made the cases that are not $i = 1$ and $n = 2$ any less difficult to notate. So we will define a function relative to indexing. First observe

$$C_{1,3}\left(\frac{C_{2,3}}{V_{2,3}} - C_{2,4}\left(\frac{C_{3,4}}{V_{3,4}V_{2,3}} - C_{3,5}(\ldots)\right)\right) = \sum_{k=2}^{N}(-1)^k \frac{C_{k,k+1}\prod_{j=1}^{k-1} C_{j,j+2}}{\prod_{j=2}^{k} V_{j,j+1}}$$

Everything else follows this structure, conditional on indexing. We can make an (ugly) generalization as follows. Let A be generic rearrangement of $\boldsymbol{X}$; i.e., $\boldsymbol{X}$ has indexing $\{1, \ldots, N\}$ and A's indexing can be any permutation of this order. Define $I_k^{\{A\}}$ as the index of $\boldsymbol{X}$ corresponding to the k-th element of A (formally: a mapping $I(k; \{A\})$). Now define

$$S(\{A\}) = \sum_{k=2}^{N}(-1)^k \frac{C_{I_k^{\{A\}}, I_{k+1}^{\{A\}}}\prod_{j=1}^{k-1} C_{I_j^{\{A\}},`}}{\prod_{j=2}^{k} V_{j,j+1}^{\{A\}}} \quad \text{with } V_{j,j+1}^{\{A\}} = \text{var}\left(X_{I_{j+1}^{\{A\}}}^{\perp_{I(1:j;\{A\})}}\right)$$

This will allow for a crawl towards completeness, burying as much of the index stumbling blocks as possible. Let P_N denote all permutations of $\{1, 2, \ldots, N\}$. Define $P_N^{i,n} \subseteq P_N$ as permutations with i, n as the first elements:

$$P_N^{i,n} = \left\{\sigma \in P_N : \sigma(1) = i \ \& \ \sigma(2) = n\right\}$$

Then we can write the earlier expression $\text{cov}(X_1, X_2^{\perp_1})$ in the general case as

$$\text{cov}(X_i, X_n^{\perp_i}) = C_{i,n} - \boldsymbol{\Sigma}_{i,n} \quad \text{with } \boldsymbol{\Sigma}_{i,n} = \sum_{\sigma \in P_N^{i,n}} S(\{\sigma\})$$

At the very beginning we started with

$$\omega_i = \frac{\text{cov}(1_{a \le \varepsilon_t}, X_i^{\perp})}{\text{var}(X_i^{\perp})}$$

And now we can write $\text{cov}(1_{a \le \varepsilon_t}, X_i^{\perp})$ compactly as

$$\text{cov}(1_{a \le \varepsilon_t}, X_i^{\perp}) = \text{cov}(1_{a \le \varepsilon_t}, X_i) - \sum_{n \ge 1 : n \ne i}^{N} \text{cov}(1_{a \le \varepsilon_t}, X_n) \cdot \left(\frac{C_{i,n} - \boldsymbol{\Sigma}_{i,n}}{\text{var}(X_n^{\perp_i})}\right)$$

We are not out of the woods yet because we skipped unpacking the variance terms. But once that has been done we will have finished "simplifying", in that arbitrarily complex regressions can be defined in terms of estimands that feature only explicit variance and covariance terms.[1]

The strategy to deal with the variance terms will be very similar and hopefully easier to digest now that we have some machinery to work with. Again to deal with the simplest case ($i = 1$) first,

$$\text{var}(X_1^{\perp}) = \text{var}(X_1) + \sum_{n=2}^{N}\sum_{m=2}^{N} \frac{\text{cov}(X_1, X_n^{\perp_1})}{V_{1,n}} \frac{\text{cov}(X_1, X_m^{\perp_1})}{V_{1,m}} C_{n,m} - \sum_{n=2}^{N} \frac{\text{cov}(X_1, X_n^{\perp_1})}{V_{1,n}} 2C_{1,n}$$

using the notation as before for V and C. Once more, we have a situation where everything will follow this pattern, less indexing. The first layer is simple to write

$$\text{var}(X_i^{\perp}) = \text{var}(X_i) + \sum_{m,n\geq 1:m,n\neq i}^{N}\sum^{N} \frac{\mathbf{\Sigma}_{i,n}}{V_{1,n}} \frac{\mathbf{\Sigma}_{i,m}}{V_{1,m}} C_{n,m} - \sum_{n\geq 1:n\neq i}^{N} \frac{\mathbf{\Sigma}_{i,n}}{V_{1,n}} 2C_{i,n}$$

The only thing remaining is to expand this definition so that it holds as terms are continually added to $\perp$ in $X_i^{\perp}$ (i.e., in the FWL regressions, some terms have already been partialed out and won't be included). To do this, we need to make the indexing of $\mathbf{\Sigma}_{i,n}$ a bit more flexible. Define

$$SV(i;\{O\}) = \text{var}(X_i^{\perp_{\{O\}}}) = \text{var}(X_i) + \sum_{m,n\geq 1:m,n\notin O}^{N}\sum^{N} \frac{\mathbf{\Sigma}^c_{\{O\},n}}{SV(n;\{O,i\})} \frac{\mathbf{\Sigma}^c_{\{O\},m}}{SV(m;\{O,i\})} C_{n,m} - \sum_{n\geq 1:n\notin O}^{N} \frac{\mathbf{\Sigma}^c_{\{O\},n}}{SV(n;\{O,i\}} 2C_{i,n}$$

where O is a set of unique integers $o \in [1,N]\backslash\{i\}$ and

$$\mathbf{\Sigma}^c_{\{O\},n} = \sum_{\sigma \in P_-^{\{O\}}} S(\{\sigma\})$$

$$\text{with } \; P_-^{\{O\}} = \left\{\sigma \in \mathbb{Z}^{N-|O|} : \sigma \cup \{O\} \in P_N \;\&\; \forall n, \nexists m \text{ s.t } \sigma(m) = O(n) \right\}.$$

Noting that for any $\sigma \in P_N$, $SV\Big(i, \big\{\sigma\backslash\{i\}\big\}\Big) = \text{var}(X_i)$, our nightmare is finally over.

Standard Errors for Generated Regressors

Recall the generated regressor functions defined in Section 3 $\left\{f_i\right\}_{i=1}^4$. For clarity, we are interested in functions of a shock ε_t that is continuously distributed on $a \in I \subset \mathbb{R}$. Each of this function has a designated "peak" c_i and a set I_i with endpoints left_i and right_i). These functions are defined in terms of the empirical CDF $F_N(\cdot)$. Specifically, for each $a \in I$

$$f_i(a) = \begin{cases} 0 & a \notin [\text{left}_i, \text{right}_i) \\ -[F_N(c_i) - F_N(\text{left}_i)]^{-1} & a \in [\text{left}_i, c) \\ [F_N(\text{right}_i) - F_N(c_i)]^{-1} & a \in (c, \text{right}_i) \end{cases}$$

with slight abuse of notation if $\text{left}_i = -\infty$. In a regression of y on $\left\{f_i\right\}_{i=1}^4$, the estimands will be defined in terms

[1]Of course, $(X'X)^{-1}X'Y$ is a better simplification under any sensible definition. Morphing "transparent" into a verb would be more appropriate

of the CDF $F(\cdot)$. Define p_{iL} as $F(c_i) - F(\text{left}_i)$ and $p_{iR} = F(\text{right}_i) - F_N(c_i)$. The estimand β_i on f_i is

$$\beta_i = \frac{\text{cov}(y, f_i)}{\text{Var}(f_i)} = \frac{\bar{y}_{iR} - \bar{y}_{iL}}{\frac{1}{p_{iL}} + \frac{1}{p_{iR}}},$$

where $\bar{y}_{iL}$ and $\bar{y}_{iR}$ are the means of y on the subsets of I_i given by p_{iL} and p_{iR}. To see this, recall f_i is mean 0. So

$$\text{cov}(y, f_i) = \mathbb{E}[y f_i] = -\frac{1}{p_{iL}}\mathbb{E}[y \cdot \mathbb{1}_{[\text{left}_i, c_i)}] + \frac{1}{p_{iR}}\mathbb{E}[y \cdot \mathbb{1}_{[c_i, \text{right}_i)}] = -\frac{1}{p_{iL}}\bar{y}_{iL} \cdot p_{iL} + \frac{1}{p_{iR}}\bar{y}_{iR} \cdot p_{iR} = \bar{y}_{iR} - \bar{y}_{iL}$$

Because this estimand is formed with respect to a generated regressor, we need to adjust the standard errors. Adjustment is done using the delta method. Differentiating

$$\frac{\partial \beta_i}{\partial p_{iL}} = \beta_i \cdot \frac{p_{iR}}{p_{iL}(p_{iL} + p_{iR})} \ \underline{\textbf{and}} \ \frac{\partial \beta_i}{\partial p_{iR}} = -\beta_i \cdot \frac{p_{iL}}{p_{iR}(p_{iL} + p_{iR})}$$

The adjustment takes the form of [2]

$$\left(\frac{\partial \beta_i}{\partial p_{iL}}\right)^2 \text{Var}(p_{iL}) + \left(\frac{\partial \beta_i}{\partial p_{iR}}\right)^2 \text{Var}(p_{iR}).$$

Using sample analogs, standard errors are the square soot of the sum of the usual Huber-White variance and

$$\frac{\hat{\beta}_i^2}{N}\left(\frac{\hat{p}_{iR}(1 - \hat{p}_{iL})}{\hat{p}_{iL}(\hat{p}_{iL} + \hat{p}_{iR})^2} + \frac{\hat{p}_{iL}(1 - \hat{p}_{iR})}{\hat{p}_{iR}(\hat{p}_{iL} + \hat{p}_{iR})^2}\right)$$

where $\widehat{\text{Var}}(p_{iL}) = \frac{\hat{p}_{iL}(1 - \hat{p}_{iL})}{N}$ (and similar for p_{iR}).

Equilibrium Model Estimation Details

For the set of draws that came out of our Metropolis- Hastings routine, I simulated data of 400 observations for each group of parameters to align with the US data sample size. Analogous control variables are included (lagged interest rates, zero lower bound, unemployment, output and interest rate variance) and plots are in terms of standard deviations to abstract away from any differences between model-simulated and US data.

	h		
	0	1	2
Big Cut	-18.8%	-5.5%	-1.7%
Big Hike	36.6%	5.5%	0.4%

Table 1: Average % Deviation from i^*, h periods after large change in i_t

[2] Note that f_i is not differentiable at c_i