This document contains brief notes for manually computing the derivatives needed to compute the AMIP and related scores for OLS and IV regression. The original versoin of zaminfluence used Python autograd to compute these derivatives. In order to avoid Python and autograd dependence for the common cases of OLS and IV, I will here derive the needed derivatives by hand. These notes comprise no more nor less than what the authors required to implement the derivatives, and may be either more or less verbose than you, the reader, require.

It suffices to do the derivations for IV, of which OLS is a special case.

$$\begin{aligned}
\epsilon_n &:= y_n - x_n^T \theta \\
G\left(\theta, w\right) &= \sum_n w_n \epsilon_n z_n \Rightarrow \\
\frac{\partial G\left(\theta, w\right)}{\partial w_n} &= w_n z_n \epsilon_n \\
\frac{\partial G\left(\theta, w\right)}{\partial \theta^T} &= -\sum_n w_n z_n x_n^T.
\end{aligned}$$

Let $X$, $Z$, and $Y$ be the matrix versions of the regressors, instruments, and responses. Let $Z_w$ denote the matrix whose $n-$th row is $w_n z_n$, and $Z_\epsilon$ denote the matrix whose $n-$th row is $w_n \left(y_n - x_n^T \hat{\theta}\right) z_n$.

It follows that

$$\frac{d\hat{\theta}}{dw^T} = \left(Z_w^T X\right)^{-1} Z_\epsilon^T.$$

It is the standard errors that require some tedious algebra.

## Matrix inverses

Recall the matrix inverse derivative formula.

$$\begin{aligned}
\frac{d}{dw_n} \left(Z_w^T X\right)^{-1} &= -\left(Z_w^T X\right)^{-1} \left(z_n x_n^T\right) \left(Z_w^T X\right)^{-1} \\
\frac{d}{dw_n} \left(X^T Z_w\right)^{-1} &= -\left(X^T Z_w\right)^{-1} \left(x_n z_n^T\right) \left(X^T Z_w\right)^{-1}.
\end{aligned}$$

## Ungrouped SEs

In the ungrouped case, the full standard error matrix is given by

$$\hat{\Sigma} = \hat{\sigma}^2 \left(Z_w^T X\right)^{-1} \left(Z_w^T Z\right) \left(Z_w^T X\right)^{-1}.$$

For the standard errors, in the case to match the output of lm, we have

$$\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{N-D}\sum_{n=1}^{N} w_n \epsilon_n^2 \Rightarrow \\
\frac{\partial \hat{\sigma}^2}{\partial \theta} &= \frac{-2}{N-D}\sum_{n=1}^{N} w_n \epsilon_n x_n.
\end{aligned}$$

For OLS, $d\hat{\sigma}^2/d\theta = 0$. Additionally, $\hat{\sigma}$ has explicit weight dependence:

$$\frac{\partial \hat{\sigma}^2}{\partial w_n} = \frac{\epsilon_n^2}{N-D}.$$

Alternatively, if we use

$$\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{\sum_n w_n - D}\sum_{n=1}^{N} w_n \epsilon_n^2 \Rightarrow \\
\frac{\partial \hat{\sigma}^2}{\partial w_n} &= \frac{\hat{\sigma}^2}{\sum_n w_{0n} - D}\left(\epsilon_n^2 - \hat{\sigma}^2\right).
\end{aligned}$$

The chain rule and the above matrix inverse formula gives

$$\begin{aligned}
\frac{d}{dw_n}\hat{\Sigma} = \frac{d\hat{\sigma}^2}{dw_n}&\left(Z_w^T X\right)^{-1}\left(Z_w^T Z\right)\left(X^T Z_w\right)^{-1} + \\
&\hat{\sigma}^2\left(Z_w^T X\right)^{-1} z_n z_n^T \left(Z_w^T X\right)^{-1} - \\
&\hat{\sigma}^2\left[\left(Z_w^T X\right)^{-1}\left(z_n x_n^T\right)\left(Z_w^T X\right)^{-1}\right]\left(Z_w^T Z\right)\left(X^T Z_w\right)^{-1} - \\
&\hat{\sigma}^2\left(Z_w^T X\right)^{-1}\left(Z_w^T Z\right)\left[\left(X^T Z_w\right)^{-1}\left(x_n z_n^T\right)\left(X^T Z_w\right)^{-1}\right].
\end{aligned}$$

Note that if

$$\begin{aligned}
A &= \left(Z_w^T X\right)^{-1}\left(Z_w^T Z\right) \\
r_{x,n} &= \left(X^T Z_w\right)^{-1} x_n \\
r_{z,n} &= \left(Z_w^T X\right)^{-1} z_n
\end{aligned}$$

then

$$\begin{aligned}
\left[\left(Z_w^T X\right)^{-1}\left(z_n x_n^T\right)\left(Z_w^T X\right)^{-1}\right]\left(Z_w^T Z\right)\left(X^T Z_w\right)^{-1} &= r_{z,n} r_{x,n}^T A^T \\
\left(Z_w^T X\right)^{-1}\left(Z_w^T Z\right)\left[\left(X^T Z_w\right)^{-1}\left(x_n z_n^T\right)\left(X^T Z_w\right)^{-1}\right] &= A r_{x,n} r_{z,n}^T \\
\left(Z_w^T X\right)^{-1} z_n z_n^T \left(Z_w^T X\right)^{-1} &= r_{z,n} r_{z,n}^T.
\end{aligned}$$

We typically only care about the diagonal, which is more convenient for matrix-valued expressions. By writing

$$\begin{aligned}
R_x &:= \left(X^T Z_w\right)^{-1} X^T \\
R_z &:= \left(Z_w^T X\right)^{-1} Z^T,
\end{aligned}$$

we can express the derivative using the element-wise (Hadamard) product $\odot$ as

$$\begin{aligned}\frac{d}{dw^T}\mathrm{diag}\left(\hat{\Sigma}\right) = \frac{d\hat{\sigma}^2}{dw_n}\mathrm{diag}\left(\left(Z_w^T X\right)^{-1}\left(Z_w^T Z\right)\left(X^T Z_w\right)^{-1}\right) + \\ \hat{\sigma}^2\left(R_z \odot R_z\right) - 2\hat{\sigma}^2\left(\left(AR_x\right)\odot R_z\right).\end{aligned}$$

Incidentally,

$$\begin{aligned}AR_x &= \left(Z_w^T X\right)^{-1}\left(Z_w^T Z\right)\left(X^T Z_w\right)^{-1} X^T \\ &= \left(\left(Z_w^T X\right)^{-1}\left(Z_w^T Z\right)\left(X^T Z_w\right)^{-1}\right) X^T,\end{aligned}$$

where the first term is just the “sandwich matrix.” Noticing this could save an inverse as well as connect the result to the grouped standard errors.

## Grouped SEs

In the grouped case, the full standard error matrix is given by the grouped score function. Let $g\left(n\right)$ denote the group to which observation $n$ belongs, which is in $1\ldots G$, and let

$$\begin{aligned}\gamma_g &:= \sum_{n:g(n)=g} \epsilon_n w_n z_n \\ S_g &:= \gamma_g - \frac{1}{G}\sum_{g=1}^{N}\gamma_g \\ &= \sum_{n:g(n)=g} \epsilon_n w_n z_n - \frac{1}{G}\sum_{n=1}^{N}\epsilon_n w_n z_n \\ S &:= \begin{pmatrix} S_1^T \\ \vdots \\ S_G^T \end{pmatrix} \\ \hat{V} &:= \frac{1}{G} S^T S \\ \hat{\Sigma} &= G\left(Z_w^T X\right)^{-1}\hat{V}\left(X^T Z_w\right)^{-1}.\end{aligned}$$

The dependence on $\left(Z_w^T X\right)^{-1}$ is the same as in the un-grouped case. Let us consider derivatives of $\hat{V}$. First,

$$\frac{\partial}{\partial w_n}\hat{V} = \frac{1}{G}S^T\frac{\partial S}{\partial w_n} + \left(\frac{1}{G}S^T\frac{\partial S}{\partial w_n}\right)^T$$

$$\frac{\partial}{\partial w_n}\gamma_{g(n)} = \epsilon_n z_n, \quad 0 \text{ otherwise}$$

$$\frac{\partial S}{\partial w_n} = \left(\begin{array}{c} -\frac{1}{G}\sum_{g=1}^N \frac{\partial}{\partial w_n}\gamma_g \\ \vdots \\ \frac{\partial}{\partial w_n}\gamma_g - \frac{1}{G}\sum_{g=1}^N \frac{\partial}{\partial w_n}\gamma_g \\ \vdots \\ -\frac{1}{G}\sum_{g=1}^N \frac{\partial}{\partial w_n}\gamma_g \end{array}\right) = \left(\begin{array}{c} -\frac{1}{G} \\ \vdots \\ 1-\frac{1}{G} \\ \vdots \\ -\frac{1}{G} \end{array}\right)\epsilon_n z_n^T,$$

where the entry containing $\frac{\partial}{\partial w_n}\gamma_g$ is in the group $g(n)$ to which observation $n$ belongs. Note that $\left(\frac{1}{G},\ldots,\frac{1}{G}\right)S = 0$ since the entries of $S$ are already demeaned. So

$$\frac{1}{G}S^T\frac{\partial S}{\partial w_n} = \frac{1}{G}S_{g(n)}z_n^T\epsilon_n.$$

Note that

$$\left(Z_w^T X\right)^{-1}\frac{\partial \hat{V}}{\partial w_n}\left(X^T Z_w\right)^{-1} = \left(Z_w^T X\right)^{-1}\frac{1}{G}S^T\frac{\partial S}{\partial w_n}\left(X^T Z_w\right)^{-1} + \left(Z_w^T X\right)^{-1}\left(\frac{1}{G}S^T\frac{\partial S}{\partial w_n}\right)^T\left(X^T Z_w\right)^{-1}.$$

The first term is thus given by

$$\left(Z_w^T X\right)^{-1}\frac{1}{G}S^T\frac{\partial S}{\partial w_n}\left(X^T Z_w\right)^{-1} = \frac{1}{G}\left(Z_w^T X\right)^{-1}S_g z_n^T\left(X^T Z_w\right)^{-1}\epsilon_n.$$

The diagonal is given by

$$\text{diag}\left(\left(Z_w^T X\right)^{-1}\frac{1}{G}S^T\frac{\partial S}{\partial w_n}\left(X^T Z_w\right)^{-1}\right) = \frac{1}{G}\left(\left(Z_w^T X\right)^{-1}S_g\right)\odot\left(\left(Z_w^T X\right)^{-1}z_n\epsilon_n\right).$$

By symmetry, the second term is the same, giving

$$\text{diag}\left(\left(Z_w^T X\right)^{-1}\frac{\partial \hat{V}}{\partial w_n}\left(X^T Z_w\right)^{-1}\right) = 2\frac{1}{G}\left(\left(Z_w^T X\right)^{-1}S_g\right)\odot\left(\left(Z_w^T X\right)^{-1}z_n\epsilon_n\right).$$

Recall from above that

$$\begin{aligned}
\left(\frac{\partial}{\partial w_n}\left(Z_w^T X\right)^{-1}\right)\hat{V}\left(X^T Z_w\right)^{-1} &= -\left(Z_w^T X\right)^{-1} z_n x_n^T \left(Z_w^T X\right)^{-1}\hat{V}\left(X^T Z_w\right)^{-1} \\
\left(Z_w^T X\right)^{-1}\hat{V}\left(\frac{\partial}{\partial w_n}\left(X^T Z_w\right)^{-1}\right) &= -\left(Z_w^T X\right)^{-1}\hat{V}\left(X^T Z_w\right)^{-1} x_n z_n^T \left(X^T Z_w\right)^{-1} \\
&= \left(\left(\frac{\partial}{\partial w_n}\left(Z_w^T X\right)^{-1}\right)\hat{V}\left(X^T Z_w\right)^{-1}\right)^T.
\end{aligned}$$

Again recognizing the diagonal,

$$\begin{gathered}
\operatorname{diag}\left(\left(\frac{\partial}{\partial w_n}\left(Z_w^T X\right)^{-1}\right)\hat{V}\left(X^T Z_w\right)^{-1}\right) = \\
-\left(\left(Z_w^T X\right)^{-1} z_n\right)\odot\left(\left(Z_w^T X\right)^{-1}\hat{V}\left(X^T Z_w\right)^{-1} x_n\right).
\end{gathered}$$

Putting everything together, and letting $\tilde{S}$ be an $N \times D$ matrix which is an expanded version of $S$ so that row $n$ of $\tilde{S}$ contains row $g\left(n\right)$ of $S$,

$$\begin{gathered}
\frac{\partial}{\partial w_n}\operatorname{diag}\left(\left(Z_w^T X\right)^{-1}\hat{V}\left(X^T Z_w\right)^{-1}\right) = \\
-2\left(\left(Z_w^T X\right)^{-1} Z^T\right)\odot\left(\left(Z_w^T X\right)^{-1}\hat{V}\left(X^T Z_w\right)^{-1} X^T\right) + \\
2\frac{G-1}{G^2}\left(\left(Z_w^T X\right)^{-1}\tilde{S}^T\right)\odot\left(\left(Z_w^T X\right)^{-1}\left(Z\odot\epsilon\right)^T\right).
\end{gathered}$$

---

Now, we consider the $\hat{\theta}$ dependence. Only $\hat{V}$ depends on $\hat{\theta}$ in this case, and, as above

$$\begin{aligned}
\frac{\partial}{\partial\hat{\theta}_d}\hat{V} &= \frac{1}{G}S^T\frac{\partial S}{\partial\hat{\theta}_d} + \left(\frac{1}{G}S^T\frac{\partial S}{\partial\hat{\theta}_d}\right)^T \\
\frac{\partial\epsilon_n}{\partial\hat{\theta}} &= -x_n
\end{aligned}$$

Let's do it component-wise.

$$\begin{aligned}
\frac{\partial}{\partial\hat{\theta}_d}\gamma_{g(n)} &= \sum_{n:g(n)=g} w_n z_n\frac{\partial\epsilon_n}{\partial\hat{\theta}_d} \\
&= -\sum_{n:g(n)=g} w_n z_n x_{nd}
\end{aligned}$$

$$\frac{\partial S}{\partial\hat{\theta}_d} = -\begin{pmatrix} \vdots \\ \sum_{n:g(n)=g} w_n z_n^T x_{nd} - \frac{1}{G}\sum_{n:=1}^{N} w_n z_n^T x_{nd} \\ \vdots \end{pmatrix}.$$

From this,

$$\begin{aligned}
S^T \frac{\partial S}{\partial \hat{\theta}_d} &= \left( -\sum_g S_g \sum_{n:g(n)=g} w_n z_n^T x_{nd} + \frac{1}{G} \left( \sum_g S_g \right) \sum_{n:=1}^{N} w_n z_n^T x_{nd} \right) \\
&= -\sum_g S_g \sum_{n:g(n)=g} w_n z_n^T x_{nd}
\end{aligned}$$

because $\sum_g S_g = 0$. Let's let

$$\begin{aligned}
\Xi_{gd} &= \sum_{n:g(n)=g} \left(-w_n x_{nd}\right) z_n \in \mathbb{R}^D \\
\Xi_d &= \begin{pmatrix} \Xi_{1d}^T \\ \vdots \\ \Xi_{gd}^T \\ \vdots \\ \Xi_{Gd}^T \end{pmatrix},
\end{aligned}$$

so that

$$S^T \frac{\partial S}{\partial \hat{\theta}_d} = S^T \Xi_d$$

Now,

$$\left(Z_w^T X\right)^{-1} S^T \frac{\partial S}{\partial \hat{\theta}_d} \left(X^T Z_w\right)^{-1} = \left( \left(Z_w^T X\right)^{-1} S^T \right) \left( \left(Z_w^T X\right)^{-1} \Xi_d^T \right)^T,$$

Looking at it this way, I'm not sure how you would avoid either a for loop over the entries of $\hat{\beta}$ or construction of higher-dimensional arrays. Putting things together,

$$\operatorname{diag}\left( \frac{\partial \hat{\Sigma}}{\partial \hat{\beta}_d} \right) = \frac{2}{G} \operatorname{diag}\left( \left( \left(Z_w^T X\right)^{-1} S^T \right) \left( \left(Z_w^T X\right)^{-1} \Xi_d^T \right)^T \right).$$