Modeling Issues

Brad Jones\(^1\)

\(^1\)Department of Political Science
University of California, Davis

February 12, 2010
Non-Constant Error Variance

- Begin with variance of the errors.

By assumption, the error variance is constant:

\[ E(\epsilon' \epsilon) = \sigma^2 I \]

The error variance is not tied to the regressors. If it holds, then (as we've seen):

\[
\text{var}(\hat{\beta}) = E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'] = E\left[ (X'X)^{-1}X'\epsilon\epsilon'X(X'X)^{-1} \right] = (X'X)^{-1}X'\sigma^2 I X(X'X)^{-1} = \sigma^2 (X'X)^{-1}X'X(X'X)^{-1} = \sigma^2 (X'X)^{-1}
\]

Using the MSE:

\[
V(\hat{\beta}) = \hat{\sigma}^2 (X'X)^{-1} - 1 = \frac{\epsilon'\epsilon}{n - k - 1}(X'X)^{-1}
\]
Non-Constant Error Variance

- Begin with variance of the errors.
- By assumption, the error variance is constant:

$$E(\epsilon' \epsilon) = \sigma^2 I$$
Non-Constant Error Variance

- Begin with variance of the errors.
- By assumption, the error variance is constant:
  \[ E(\epsilon'\epsilon) = \sigma^2 I \]
- The error variance is not tied to the regressors. If it holds, then (as we’ve seen):
  \[
  \text{var} (\hat{\beta}) = E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'] \\
  = E[(X'X)^{-1}X'\epsilon\epsilon'X(X'X)^{-1}] \\
  = (X'X)^{-1}X'E[\epsilon\epsilon']X(X'X)^{-1} \\
  = (X'X)^{-1}X'(\sigma^2 I)X(X'X)^{-1} \\
  = \sigma^2 (X'X)^{-1} \quad (1)
  \]
Non-Constant Error Variance

- Begin with variance of the errors.
- By assumption, the error variance is constant:

\[ E(\epsilon'\epsilon) = \sigma^2 I \]

- The error variance is not tied to the regressors. If it holds, then (as we’ve seen):

\[
\text{var}(\hat{\beta}) = E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)']
\]
\[
= E[(X'X)^{-1}X'\epsilon\epsilon'X(X'X)^{-1}]
\]
\[
= (X'X)^{-1}X'E[\epsilon\epsilon']X(X'X)^{-1}
\]
\[
= (X'X)^{-1}X'(\sigma^2 I)X(X'X)^{-1}
\]
\[
= \sigma^2 (X'X)^{-1}
\]

1

- Using the MSE:

\[
V(\hat{\beta}) = \hat{\sigma}^2 (X'X)^{-1} = \frac{\epsilon\epsilon'}{n - k - 1}(X'X)^{-1}
\]
Non-Constant Error Variance

Suppose:

\[ E(\epsilon' \epsilon) = \phi \]
Suppose: 

\[ E(\epsilon'\epsilon) = \phi \]

Then

\[ E(\hat{\beta} - \beta)(\hat{\beta} - \beta') = (X'X)^{-1}X'\phi X(X'X)^{-1} \]
Non-Constant Error Variance

- Suppose:
  
  $$E(\epsilon'\epsilon) = \phi$$

- Then
  
  $$E(\hat{\beta} - \beta)(\hat{\beta} - \beta') = (X'X)^{-1}X'\phi X(X'X)^{-1}$$

- That is, the variance is dependent on the variance of the errors. By assumption, they are constant; if $$\sigma^2 I$$, then the variance function reduces to that on the previous slide.
Non-Constant Error Variance

- Suppose:
  \[ E(\epsilon'\epsilon) = \phi \]

- Then
  \[ E(\hat{\beta} - \beta)(\hat{\beta} - \beta') = (X'X)^{-1}X'\phi X(X'X)^{-1} \]

- That is, the variance is dependent on the variance of the errors. By assumption, they are constant; if \( \sigma^2 I \), then the variance function reduces to that on the previous slide.

- That is . . . no problem.
Non-Constant Error Variance

▶ There are several remarks worth making.

If the errors are heteroskedastic, then estimates of the standard errors around the regression coefficients will be wrong. How wrong? You will not be able to tell (nor can you tell the direction of the error). What about the regression parameters? Note also we are still assuming the errors are independent. Solutions? The issue is if $\phi \neq \sigma^2$ then we may need an alternative estimator of the variance. Weighted-Least Squares?
There are several remarks worth making.

If the errors are heteroskedastic, then estimates of the standard errors around the regression coefficients will be wrong.
Non-Constant Error Variance

- There are several remarks worth making.
- If the errors are heteroskedastic, then estimates of the standard errors around the regression coefficients will be wrong.
- How wrong? You will not be able to tell (nor can you tell the direction of the error).
Non-Constant Error Variance

- There are several remarks worth making.
- If the errors are heteroskedastic, then estimates of the standard errors around the regression coefficients will be wrong.
- How wrong? You will not be able to tell (nor can you tell the direction of the error).
- What about the regression parameters?
Non-Constant Error Variance

- There are several remarks worth making.
- If the errors are heteroskedastic, then estimates of the standard errors around the regression coefficients will be wrong.
- How wrong? You will not be able to tell (nor can you tell the direction of the error).
- What about the regression parameters?
- Note also we are *still* assuming the errors are independent.
Non-Constant Error Variance

There are several remarks worth making.

If the errors are heteroskedastic, then estimates of the standard errors around the regression coefficients will be wrong.

How wrong? You will not be able to tell (nor can you tell the direction of the error).

What about the regression parameters?

Note also we are still assuming the errors are independent.

Solutions? The issue is if \( \phi \neq \sigma^2 I \) then we may need an alternative estimator of the variance.
There are several remarks worth making.

If the errors are heteroskedastic, then estimates of the standard errors around the regression coefficients will be wrong.

How wrong? You will not be able to tell (nor can you tell the direction of the error).

What about the regression parameters?

Note also we are still assuming the errors are independent.

Solutions? The issue is if $\phi \neq \sigma^2 I$ then we may need an alternative estimator of the variance.

Weighted-Least Squares?
Non-Constant Error Variance

Note that $E(\epsilon) = 0$ but the variance is now $\sigma_i^2$ (that is, it's heteroskedastic).
Non-Constant Error Variance

- Note that $E(\epsilon) = 0$ but the variance is now $\sigma_i^2$ (that is, it's heteroskedastic).
- Suppose we identify an estimator for $\phi$ and call it $\hat{\phi}$?
Non-Constant Error Variance

- Note that $E(\epsilon) = 0$ but the variance is now $\sigma_i^2$ (that is, it’s heteroskedastic).
- Suppose we identify an estimator for $\phi$ and call it $\hat{\phi}$?
- If we could show this to be a consistent estimator of the variance of $\beta$ then we would be in luck.
Note that $E(\epsilon) = 0$ but the variance is now $\sigma_i^2$ (that is, it’s heteroskedastic).

Suppose we identify an estimator for $\phi$ and call it $\hat{\phi}$?

If we could show this to be a consistent estimator of the variance of $\beta$ then we would be in luck.

Enter Hal White (1982). White proposed the use of $\text{diag}[\hat{E}_1, \ldots, \hat{E}_n]$, where $E_i$ is the residual for the $i$th observation.
Non-Constant Error Variance

- Note that $E(\epsilon) = 0$ but the variance is now $\sigma_i^2$ (that is, it’s heteroskedastic).
- Suppose we identify an estimator for $\phi$ and call it $\hat{\phi}$?
- If we could show this to be a consistent estimator of the variance of $\beta$ then we would be in luck.
- Enter Hal White (1982). White proposed the use of $\text{diag}[\hat{E}_1, \ldots, \hat{E}_n]$, where $E_i$ is the residual for the $i$th observation.
- This gives rise to:

\[
\tilde{V}(\hat{\beta}) = (X'X)^{-1}X'\text{diag}[e^2i]X(X'X)^{-1} \\
= (X'X)^{-1}X'\hat{\phi}X(X'X)^{-1}
\]  

(3)
The entries on the main diagonal are the estimated squared standard errors of the regression coefficients.
Non-Constant Error Variance

- The entries on the main diagonal are the estimated squared standard errors of the regression coefficients.
- Use of this form produces what are sometimes called “White corrected standard errors.”
Non-Constant Error Variance

- The entries on the main diagonal are the estimated squared standard errors of the regression coefficients.
- Use of this form produces what are sometimes called “White corrected standard errors.”
- Specifically, this form is a “heteroskedastic consistent” estimator of the variance.
Non-Constant Error Variance

- The entries on the main diagonal are the estimated squared standard errors of the regression coefficients.
- Use of this form produces what are sometimes called “White corrected standard errors.”
- Specifically, this form is a “heteroskedastic consistent” estimator of the variance.
- This is also known as a “sandwich estimator” and is sometimes called a Huber-White sandwich estimator.
Non-Constant Error Variance

- The entries on the main diagonal are the estimated squared standard errors of the regression coefficients.
- Use of this form produces what are sometimes called “White corrected standard errors.”
- Specifically, this form is a “heteroskedastic consistent” estimator of the variance.
- This is also known as a “sandwich estimator” and is sometimes called a Huber-White sandwich estimator.
- “Huber” because Huber’s 1967 paper anticipated White’s ground-breaking paper.
Non-Constant Error Variance

- Implementation in R can be done through coeftest
Non-Constant Error Variance

- Implementation in R can be done through `coeftest`
- ```r
  coeftest(mod.duncan, vcov = vcovHC(mod.duncan, type = "HC0"))
```
Non-Constant Error Variance

- Implementation in R can be done through `coeftest`
  - `coeftest(mod.duncan, vcov = vcovHC(mod.duncan, type = "HC0"))`
- HC0 returns White’s standard errors.

Fox discusses Long and Ervin’s estimator:

\[
\tilde{V}(\hat{\beta}) = \text{diag}\left[\frac{\hat{E}_1}{(1 - h_{ii})^2}, \ldots, \frac{\hat{E}_n}{(1 - h_{ii})^2}\right]
\]

These are obtained by HC3 in the above statement.

Extensions: suppose there is within-unit correlation among the observations.
Non-Constant Error Variance

- Implementation in R can be done through `coeftest`
- `coeftest(mod.duncan, vcov = vcovHC(mod.duncan, type = "HC0"))`
- HC0 returns White’s standard errors.
- Fox discusses Long and Ervin’s estimator:
  \[
  \tilde{V}(\hat{\beta}) = \text{diag}[\hat{E}_1/(1 - h_i)^2, \ldots, \hat{E}_n/(1 - h_i)^2]
  \]
Non-Constant Error Variance

- Implementation in R can be done through `coef_test`
- `coef_test(mod.duncan, vcov = vcovHC(mod.duncan, type = "HC0"))`
- HC0 returns White’s standard errors.
- Fox discusses Long and Ervin’s estimator:
  \[ \tilde{V}(\hat{\beta}) = \text{diag}[\hat{E}_1/(1 - h_i)^2, \ldots, \hat{E}_n/(1 - h_i)^2] \]
- These are obtained by HC3 in the above statement.
Non-Constant Error Variance

- Implementation in R can be done through `coeftest`
  
  ```
  coeftest(mod.duncan, vcov = vcovHC(mod.duncan, type = "HC0"))
  ```

- HC0 returns White’s standard errors.

- Fox discusses Long and Ervin’s estimator:
  
  $$
  \tilde{V}(\hat{\beta}) = \text{diag}[\hat{\mathcal{E}}_1/(1 - h_i)^2, \ldots, \hat{\mathcal{E}}_n/(1 - h_i)^2]
  $$

- These are obtained by HC3 in the above statement.

- Extensions: suppose there is within-unit correlation among the observations.
Further Motivation

- Imagine an analyst has survey data on an outcome variable of interest as well as data on covariates.
Further Motivation

- Imagine an analyst has survey data on an outcome variable of interest as well as data on covariates.
- Suppose these data are collected across countries.
Imagine an analyst has survey data on an outcome variable of interest as well as data on covariates. Suppose these data are collected across countries. In this setting, the country would denote the “j unit” and the respondent would denote the “i unit.”
Further Motivation

- Imagine an analyst has survey data on an outcome variable of interest as well as data on covariates.
- Suppose these data are collected across countries.
- In this setting, the country would denote the “j unit” and the respondent would denote the “i unit.”
- If the researcher is interested in modeling the relationship between a response variable $y_i$ and some individual-level covariate, $x_i$, a garden-variety regression model could be estimated,

$$ y_i = \beta_0 + \beta_1 x_i + \epsilon_i. $$

(4)
Further Motivation

- If estimated using all $n$ cases, the researcher is explicitly pooling the data.
Further Motivation

- If estimated using all $n$ cases, the researcher is explicitly \textit{pooling} the data.
- If the data were cross-sectional, as in the case of examining a single administration of a survey, this modeling strategy is equivalent to stacking each survey from country $j$. 
Further Motivation

- If estimated using all $n$ cases, the researcher is explicitly pooling the data.
- If the data were cross-sectional, as in the case of examining a single administration of a survey, this modeling strategy is equivalent to stacking each survey from country $j$.
- Optimistically, this approach proceeds as if cross-unit heterogeneity is nonexistent; each of the $j$ units are homogenous.
Further Motivation

- If estimated using all \( n \) cases, the researcher is explicitly *pooling* the data.
- If the data were cross-sectional, as in the case of examining a single administration of a survey, this modeling strategy is equivalent to stacking each survey from country \( j \).
- Optimistically, this approach proceeds as if cross-unit heterogeneity is nonexistent; each of the \( j \) units are homogenous.
- In the context of (4), unit heterogeneity may induce *nonspherical* disturbances.
Further Motivation

If estimated using all \( n \) cases, the researcher is explicitly *pooling* the data.

If the data were cross-sectional, as in the case of examining a single administration of a survey, this modeling strategy is equivalent to stacking each survey from country \( j \).

Optimistically, this approach proceeds as if cross-unit heterogeneity is nonexistent; each of the \( j \) units are homogenous.

In the context of (4), unit heterogeneity may induce *nonspherical* disturbances.

Heteroskedasticity may arise because the \( i \) observations in the \( j \) units are subject to different political conditions or simply because measurement error in \( x_i \) varies across \( j \).
Further Motivation

- In either case, the model is agnostic with respect to heterogeneity and so $\text{var}(\epsilon)$ is no longer a constant, $\sigma^2$, but a variable, $\sigma_i^2$. 
Further Motivation

- In either case, the model is agnostic with respect to heterogeneity and so \( \text{var}(\epsilon) \) is no longer a constant, \( \sigma^2 \), but a variable, \( \sigma_i^2 \).

- **Autocorrelation** may arise in this kind of model because the \( i \) respondents in unit \( j \) may be more alike—hence correlated—with each other than, say, the \( i \) respondents in some other unit, \( k \).
Further Motivation

- In either case, the model is agnostic with respect to heterogeneity and so $\text{var}(\epsilon)$ is no longer a constant, $\sigma^2$, but a variable, $\sigma_i^2$.
- **Autocorrelation** may arise in this kind of model because the $i$ respondents in unit $j$ may be more alike—hence correlated—than, say, the $i$ respondents in some other unit, $k$.
- This kind of correlation is sometimes called the intraclass correlation.
Further Motivation

- In either case, the model is agnostic with respect to heterogeneity and so $\text{var}(\epsilon)$ is no longer a constant, $\sigma^2$, but a variable, $\sigma^2_i$.

- **Autocorrelation** may arise in this kind of model because the $i$ respondents in unit $j$ may be more alike—hence correlated—with each other than, say, the $i$ respondents in some other unit, $k$.

- This kind of correlation is sometimes called the intraclass correlation.

- The usual assumption of zero covariance between disturbances, conditional on the covariates, may not hold.
Further Motivation

- In either case, the model is agnostic with respect to heterogeneity and so $\text{var}(\epsilon)$ is no longer a constant, $\sigma^2$, but a variable, $\sigma_i^2$.

- *Autocorrelation* may arise in this kind of model because the $i$ respondents in unit $j$ may be more alike—hence correlated—with each other than, say, the $i$ respondents in some other unit, $k$.

- This kind of correlation is sometimes called the intraclass correlation.

- The usual assumption of zero covariance between disturbances, conditional on the covariates, may not hold.

- Since the assumption of spherical disturbances is a conditional one, these problems can be mitigated by better specification of (4), for example, inclusion of covariates thought to “explain” or account for unit-wise heterogeneity.
Further Motivation

- Absent this, blissful estimation of (4) in the presence of the kinds of problems commonly yields inefficient and inconsistent standard errors.
Further Motivation

- Absent this, blissful estimation of (4) in the presence of the kinds of problems commonly yields inefficient and inconsistent standard errors.

- Moreover, because the intraclass correlation will usually be positive (Hox 2002, Kreft and De Leeuw 1998), standard errors will usually be attenuated, thus increasing the chances for a Type-I error (Barcikowski 1981).
Further Motivation

- Absent this, blissful estimation of (4) in the presence of the kinds of problems commonly yields inefficient and inconsistent standard errors.

- Moreover, because the intraclass correlation will usually be positive (Hox 2002, Kreft and De Leeuw 1998), standard errors will usually be attenuated, thus increasing the chances for a Type-I error (Barcikowski 1981).

- White’s heteroskedastic consistent standard errors (White 1980) is a common solution for part of the problem.

\[ \hat{V}_{\text{White}} = \left( X'X \right)^{-1} \left[ N \sum_{i=1}^{N} (e_i x_i) (e_i x_i)' \right] \left( X'X \right)^{-1}. \] (5)
Further Motivation

Absent this, blissful estimation of (4) in the presence of the kinds of problems commonly yields inefficient and inconsistent standard errors.

Moreover, because the intraclass correlation will usually be positive (Hox 2002, Kreft and De Leeuw 1998), standard errors will usually be attenuated, thus increasing the chances for a Type-I error (Barcikowski 1981).

White’s heteroskedastic consistent standard errors (White 1980) is a common solution for part of the problem.

White’s solution does not account for clustering:

\[
\hat{V}_W = (X'X)^{-1} \left[ \sum_{i=1}^{N} (e_i x_i)' (e_i x_i) \right] (X'X)^{-1}.
\]
Further Motivation

- Under (5), each observation contributes one variance term and so no adjustment is made to account for the grouping of observations within $j$ units.
Further Motivation

- Under (5), each observation contributes one variance term and so no adjustment is made to account for the grouping of observations within $j$ units.

- The general form of the variance estimator for clustered data is given by

$$
\hat{\mathbf{V}}_C = (\mathbf{X}'\mathbf{X})^{-1} \left[ \sum_{j=1}^{n_c} \left\{ \left( \sum_{i=1}^{n_j} e_i \mathbf{x}_i \right)' \left( \sum_{i=1}^{n_j} e_i \mathbf{x}_i \right) \right\} \right] (\mathbf{X}'\mathbf{X})^{-1},
$$

where $n_c$ corresponds to the number of clusters and $n_j$ corresponds to the number of $i$-cases within unit $j$. 
Further Motivation

- Under (5), each observation contributes one variance term and so no adjustment is made to account for the grouping of observations within $j$ units.
- The general form of the variance estimator for clustered data is given by

$$
\hat{V}_C = (X'X)^{-1} \left[ \sum_{j=1}^{n_c} \left\{ \left( \sum_{i=1}^{n_j} e_i x_i \right)' \left( \sum_{i=1}^{n_j} e_i x_i \right) \right\} \right] (X'X)^{-1},
$$

where $n_c$ corresponds to the number of clusters and $n_j$ corresponds to the number of $i$-cases within unit $j$.

- It is evident from (6) that clustering is explicitly accounted for because the cross-products of the residuals and regressors are computed first within the $j$th cluster and then summed over the $n_j$ clusters (see Franzese 2005 for further discussion of this).
Further Motivation

- Under (5), each observation contributes one variance term and so no adjustment is made to account for the grouping of observations within $j$ units.

- The general form of the variance estimator for clustered data is given by

\[ \hat{V}_C = (X'X)^{-1} \left[ \sum_{j=1}^{n_c} \left\{ \left( \sum_{i=1}^{n_j} e_i x_i \right) \left( \sum_{i=1}^{n_j} e_i x_i \right)' \right\} \right] (X'X)^{-1}, \]

(6)

where $n_c$ corresponds to the number of clusters and $n_j$ corresponds to the number of $i$-cases within unit $j$.

- It is evident from (6) that clustering is explicitly accounted for because the cross-products of the residuals and regressors are computed first within the $j$th cluster and then summed over the $n_j$ clusters (see Franzese 2005 for further discussion of this).
Turn attention now to collinearity.
Collinearity

- Turn attention now to collinearity.
- What does regression say about this?

No perfect collinearity.

With perfect collinearity, \( X'X - I \) does not exist (why?).

We will see that even in the face of highly collinear variables, the coefficients are unbiased... it's just that the standard errors will get large.
Collinearity

- Turn attention now to collinearity.
- What does regression say about this?
- No *perfect* collinearity.
Collinearity

- Turn attention now to collinearity.
- What does regression say about this?
- No \emph{perfect} collinearity.
- With perfect collinearity, \((X'X)^{-1}\) does not exist (why?)
Turn attention now to collinearity.

What does regression say about this?

No *perfect* collinearity.

With perfect collinearity, \((X'X)^{-1}\) does not exist (why?)

We will see that even in the face of highly collinear variables, the coefficients are unbiased . . . it’s just that the standard errors will get large.
Collinearity

- The “problem”.

\[ Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon \]

Rewriting:

\[ Y = \gamma_0 + \gamma_1 X_1 + \gamma_2 (X_1 - X_2) + \epsilon \]

\[ = \gamma_0 + \gamma_1 X_1 - \gamma_2 X_2 + \epsilon \]

\[ = \gamma_0 + (\gamma_1 + \gamma_2) X_1 - \gamma_2 X_2 + \epsilon \] (7)

Implications?
Collinearity

- The “problem”.
- The model:

\[ Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon \]
Collinearity

- The “problem”.
- The model:
  \[ Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon \]
- Rewriting:
  \[ Y = \gamma_0 + \gamma_1 X_1 + \gamma_2 (X_1 - X_2) + \epsilon \]
  \[ = \gamma_0 + \gamma_1 X_1 + \gamma_2 X_1 - \gamma_2 X_2 + \epsilon \]
  \[ = \gamma_0 + (\gamma_1 + \gamma_2) X_1 - \gamma_2 X_2 + \epsilon \] (7)
Collinearity

- The “problem”.
- The model:
  \[ Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon \]
- Rewriting:
  \[ Y = \gamma_0 + \gamma_1 X_1 + \gamma_2 (X_1 - X_2) + \epsilon \]
  \[ = \gamma_0 + \gamma_1 X_1 + \gamma_2 X_1 - \gamma_2 X_2 + \epsilon \]
  \[ = \gamma_0 + (\gamma_1 + \gamma_2) X_1 - \gamma_2 X_2 + \epsilon \]
  \[ (7) \]
- Implications?
Collinearity

▶ How to check?

- How to check?

  - Why not start by inspecting the simple (linear correlations)?
  - High correlations might suggest you may have a problem.
  - Inspect auxiliary regressions as well. (Recall the standard error of $\hat{\beta}$.)
  - Compute the VIF: $VIF = \frac{1}{1 - R^2_j}$.
Collinearity

- How to check?
- Why not start by inspecting the simple (linear correlations)?

- Inspect auxiliary regressions as well. (Recall the standard error $\hat{\beta}$.)

- Compute the VIF:
  $$VIF = \frac{1}{1 - R^2_j}.$$
Collinearity

- How to check?
- Why not start by inspecting the simple (linear correlations)?
- High correlations might suggest you may have a problem.
Collinearity

- How to check?
- Why not start by inspecting the simple (linear correlations)?
- High correlations might suggest you *may* have a problem.
- Inspect auxiliary regressions as well. (Recall the standard error of $\hat{\beta}$.)
How to check?
Why not start by inspecting the simple (linear correlations)?
High correlations might suggest you may have a problem.
Inspect auxiliary regressions as well. (Recall the standard error of \( \hat{\beta} \).
Compute the VIF:

\[
VIF = \frac{1}{1 - R_j^2}
\]
Collinearity

▶ Why VIF?

Recall:

$V(\hat{\beta}) = \hat{\sigma}^2 (n-1) \text{var}(X_j) \times (1 - R^2_j)$

The second term is the variance inflation factor and it shows exactly what is going on: as the correlation between $X_k$ and the other $X_j$ increases, the variance about the estimator also increases.

This is the “problem” with collinearity.

Solutions?
Collinearity

- Why VIF?
- Recall:

\[ V(\hat{\beta}) = \frac{\sigma^2}{(n - 1)\text{var}(X_j)} \times \frac{1}{1 - R_j^2} \]
Collinearity

Why VIF?

Recall:

\[ V(\hat{\beta}) = \frac{\hat{\sigma}^2}{(n-1)\text{var}(X_j)} \times \frac{1}{1 - R_j^2} \]

The second term is the variance inflation factor and it shows exactly what is going on: as the correlation between \( X_k \) and the other \( X_j \) increases, the variance about the estimator also increases.
Collinearity

Why VIF?

Recall:

\[ V(\hat{\beta}) = \frac{\hat{\sigma}^2}{(n - 1)\text{var}(X_j)} \times \frac{1}{1 - R_j^2} \]

The second term is the variance inflation factor and it shows exactly what is going on: as the correlation between \( X_k \) and the other \( X_j \) increases, the variance about the estimator also increases.

This is the “problem” with collinearity.
Collinearity

➤ Why VIF?

➤ Recall:

\[ V(\hat{\beta}) = \frac{\hat{\sigma}^2}{(n - 1)\text{var}(X_j)} \times \frac{1}{1 - R_j^2} \]

➤ The second term is the variance inflation factor and it shows exactly what is going on: as the correlation between \( X_k \) and the other \( X_j \) increases, the variance about the estimator also increases.

➤ This is the “problem” with collinearity.

➤ Solutions?