DISSERTATION PART 2

Model Interpretation with Correlated and/or Collinear Explanatory Variables

Ella Foster-Molina

Abstract

Correlated variables can obscure the interpretation of a regression model. One well understood confounder is an omitted variable. Yet including all correlated variables can also obscure the interpretation of the model. With sufficiently high correlation, the independent effect of one variable can both be hard to detect and have minimal substantive meaning. I examine cases where the combined impact of two highly correlated, jointly statistically significant variables is more meaningful than the independent impact of either. I demonstrate how to use principal component analysis to interpret their joint impact. One benefit of this approach is that it allows multicollinear variables to be interpreted. This is an extreme case of correlated variables, where the independent impacts of the variables are obscured because their correlation inflates standard errors and reduces statistical significance. The interpretation of the effect of multicollinear variables has traditionally been avoided, but can be done with either principal component analysis or a careful interpretation of omitted variable models. I use the correlation between education and income to demonstrate three cases. In the first, the independent effect of each variable is statistically significant despite being highly correlated. This is the best case scenario for a scholar interested in understanding the impact of education and income, or any other highly correlated variables. In the second case, exactly one of the collinear variables is statistically significant. In the third, a joint significance test rejects the null hypothesis that both coefficients are zero, but multicollinearity obscures the independent impact of each. I then present a guideline for interpreting models with highly correlated independent variables. I demonstrate a method to visualize the impact of multicollinearity and omitted variables. I apply this visualization and method to a series of examples that highlight the confounding and occasionally countervailing effects of education and income in range of political outcomes.

Contents

1 INTRODUCTION

The nature of research into class relations and politics often involves explanatory variables that are highly correlated with each other. Some examples include:

- partisanship ~ income + education
- educational outcomes ~ race + poverty
- turnout ~ unemployment + education

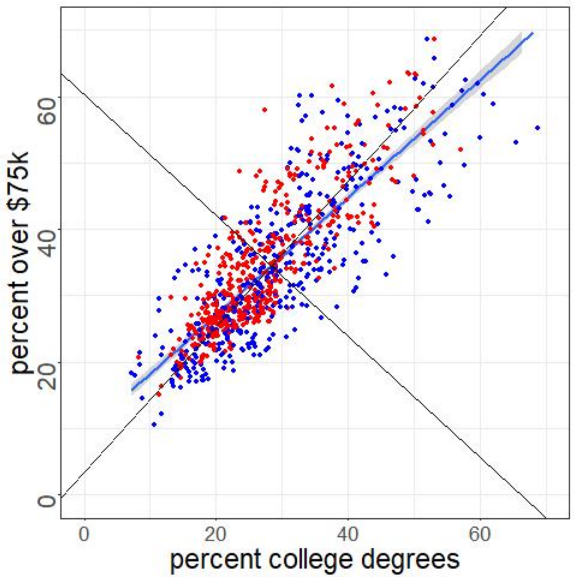
The goal is to interpret both the statistical and substantive effect of each correlated variable. Yet highly correlated variables can interfere with this goal unless careful attention is paid to model selection and interpretation. The method I present in this section of the dissertation allows researchers to test hypotheses that are more relevant to the observed data and apparent data generating process.

Consider the correlation between Congressional districts with many highly educated constituents and those with many affluent constituents, shown in Figure 1. Each point in the scatter plot is one Congressional district. The districts are color coded red if they are represented by a Republican, and blue if they are represented by a Democrat. The regression line

$$\text{income} \sim \beta_{education}{}^{*}\ \text{education} + \text{c}$$

is plotted in blue. The first principal component is the black line with a positive slope. This reflects the length along with the greatest change in education and income jointly occurs. The secondary component, the black line with a negative slope, is perpendicular to the first. With two variables, only two components are available. The equations for these lines are as follows:

Figure 1: Correlation between District Income and Education 2013-2014



$$\text{income} = 60.30586 - 0.9119 * \text{education}$$

$$\text{income} = 3.419412 + 1.0966616 * \text{education}.$$

Principal component analysis involves a change in coordinate systems. The original axes, income and education, are transformed to reflect the coordinate system defined by the two principal components. In this case, I define the first component to be socioeconomic status (SES), which reflects the joint change in income and education that will be defined in Equation 52. I call the secondary component orthogonal SES. These are calculated as:

$$\text{SES} = -0.4456 + 0.7389 * \text{income} + 0.6738 * \text{education} \tag{1}$$

$$\text{orthogonal SES} = 2.304 - 0.6738 * \text{income} + 0.7389 * \text{education}$$

Justification: Multivariate regressions are traditionally interpreted in terms of the independent impact of each variable in the regression. Yet when two explanatory variables are correlated, there is a limit to what can be interpreted based on the independent impact of one of the correlated variables. So long as the two explanatory variables are correlated, their principal components expand what can be tested and interpreted about each of the original two explanatory variables.

Analyses of regression results are based on comparing the expected difference in the dependent variable based on changes in the explanatory variable. Consider the example of district education and income as explanatory variables, shown in Figure 1. Choose any district at the low end of the income spectrum. There will be limited options for selecting a district with similar levels of education yet higher affluence. However, if I try to find a district to compare it with along the direction of the first principal component, I can find many districts to compare it to across almost the full range of affluence that exists in the observed data. The education level will necessarily change as well, but that reflects the nature of the data.

In this case, it reflects the nature of the world. There are no districts where 60% of their residents earn over $75,000, but only 20% have a college education. Moving in the direction of the first principal component reveals much about the impact of income on the dependent variable. This information is different, yet complimentary, to the information revealed by increasing income while education is held constant. In addition, the estimation of the coefficient for the first principal component will be more precise because more data is available to be leveraged for the purposes of estimating the coefficient.

Method: I propose that the joint impact of correlated variables is at least as meaningful as the independent impact of each. I introduce a set of guidelines for model interpretation when there is joint statistical significance between two highly correlated variables, broken down into three cases. In each case the coefficients on the principal components and the independent effects of each variable should be estimated and interpreted. Case 1 is the best case scenario, where the large standard errors on the collinear variables is not so high that the model fails to reject the null hypothesis for the independent impact of both collinear variables. In this case, the independent impact of the correlated variables can be understood. Yet because the variables are correlated, the principal component direction is also substantively meaningful. It will specify the potential for each variable to confound the other, and will allow the researcher to examine changes in the dependent variable based on changes in each correlated variable along almost the entire range of observed values. Case 2 is only slightly worse, as one variable is statistically significant but not the other. In this case, the traditional method has been to drop the statistically insignificant variable. This method necessarily reflects the joint impact of both the statistically insignificant and statistically significant variable, yet the joint impact is hard to specify. The principal component direction provides a more concrete interpretation for the joint impact of both variables. The last case is the trickiest, as the variables are jointly significant but neither are individually significant. This indicates that there is not sufficient data to recover the independent impact of either and that the most meaningful change to examine is one that occurs as the collinear variables move in tandem.

Benefits:

Clear definition of the joint impact: The method I propose provides a clearly visualized and mathematically specified combination of the two correlated explanatory variables. The first principal component defines a combination of the two variables that reflects the changes that occur along the heart of the observed data. This will reveal the potential for the two variables to confound each other if not properly interpreted.

Statistical significance: The effect of the joint impact of the two variables, which will also provide information about the effect of each original variable, can be tested using traditional hypothesis tests. This can be done using the same model used to understand the substantive meaning of changes in one original variable, which will coincide with changes in the correlated variable.

Interpretation of size of effect on the dependent variable: Using the principal component expands the information that can be uncovered with correlated variables. The size of the change in the dependent variable can be calculated based on the change in one of the correlated variables, so long as the other correlated variable is allowed to change concurrently in the manner specified by the principal component. Much larger changes in the explanatory variables can be accounted for than when each original variable is interpreted alone.

Easily implemented This method is easily implemented. The requirements are to plot the two correlated variables against each other to visualize their joint impact, then calculate the principal components of the two variables. The remaining implementation is the standard application of a regression, in which the original variables are replaced by their principal components.

Interpret two multicollinear variables: Multicollinearity can lead to statistically insignificant results for the original correlated variables. The principal component method

- will reveal the joint impact of the two variables if that combined impact is statistically significant even in the face of high multicollinearity.
- can be applied without collecting additional data.
- does not induce biased coefficients so long as all components are included in the model.

2 EXISTING APPROACHES to CORRELATED VARIABLES

The statistical and inferential dangers of multicollinearity have been well examined (Belsley, 1991; Montgomery, Peck, & Vining, 2012). Multicollinearity, also called collinearity, occurs when one variable can be linearly predicted by another set of variables with a substantial degree of accuracy. When all multicollinear variables are included in a model, the standard errors of the estimated coefficients for the collinear variables increase as the collinearity increases. This can result in situations where the joint effect of the collinear variables is statistically significant, but the individual coefficients are not.

The method I propose to resolve this situation is not common, and can be usefully applied to any pair of correlated variables regardless of whether the individual coefficients are statistically significant. Extensive searches of the literature have not uncovered a similar approach. Some common methods for handling multicollinear data include: ridge regression, LASSO, clustering, partial least squares regression, omitting variables, principal component regression, residual regression, and structural equation modeling. Each of these has value. My principal component[1] approach offers a number of benefits:

No bias Methods such as ridge regression and LASSO induce biased coefficients because the resulting estimators are biased. Because of this bias, computing standard errors is not recommended (Park & Casella, 2008). This makes it difficult to perform hypothesis tests, which are quite natural to perform using my principal components method. Omitting variables, as is done in clustering, induces bias due to a misspecification of which coefficient is being calculated: the one from the model without the omitted variables or the one in which variables are omitted. My principal components method does not induce bias, either in terms of biased estimators or omitted variables.

Clearly defined joint impact Methods such as clustering, omitting variables, principal component regression, residual regression, and others are often used to interpret the independent impact of individual variables. Effectively, this induces bias due to omitting information from the model and can confound the interpretation of the coefficients. I clearly define the joint impact required by the principal component, highlighting the potential for confounding while not succumbing to it.

Visualization My method allows the substantive effect of the change in the direction of the principal components to be clearly depicted, as I do in Figures 15. This is not typically possible using other methods.

Ease of implementation and interpretation The traditional approach to principal component regression is to find the principal components of the entire set of explanatory variables, then exclude those that do not provide much

[1] Principal components have been widely applied, but for distinct purposes such as prediction or measuring a latent variable that is theoretically relevant.

explanatory power to the model (Greene, 2003). My principal component is created from two correlated of explanatory variables in the regression. This allows their combination in the principal component to be clearly defined. Not only do I use a subset of explanatory variables, but I keep all resulting principal components in order to avoid biasing the remaining variables by omitting information.

I provide answers to the following commonly stated opinions regarding multicollinear data:

1. "...use of the procedures (for principal components) would hardly be necessary when the beta vector could be estimated directly by classical methods" (Massy, 1965). I show that principal components can be useful even when the beta vector can be directly estimated with statistical significance. They add more information about how each correlated variable affects the dependent variable because changes can be estimated in the dependent variable along nearly the full range of the values of the correlated explanatory variables.

2. Multicollinearity can cause problems that are "severe and sometimes crippling" (O'Brien, 2007). As we see in Example 1, my principal components approach can provide statistically significant and substantively meaningful results even in the face of high multicollinearity.

3. "Multicollinearity can create inaccurate estimates of the regression coefficients, inflate the standard errors of the regression coefficients, deflate the partial t-tests for the regression coefficients, give false, non-significant, p-values, and degrade the predictability of the model (and that's just for starters)." (*Ridge Regression*, n.d.) The visualizations highlight the effect of multicollinearity.

They show why two correlated variables would not have statistically significant effects in a multivariate regression. Namely, there is not enough data once the other correlated variable is held constant, and the resulting variation in that data is low.

4. "If a model is to be retained in all its complexity, solution of the multicollinearity problem requires an augmentation of existing data to include additional information" (Farrar & Glauber, 1967). See also Arceneaux and Huber (2007). I show that multicollinear variables can be meaningfully interpreted by using a principal component approach instead of adding new data, although of course adding new data is always valuable.

5. Existing approaches use principal components primarily to interpret the concept they were designed to measure (O'Brien, 2007). For the case with two correlated variables, I demonstrate a clear definition for the meaning of a principal component in terms of the original two variables instead of just in terms of an abstract concept of socioeconomic status.[2] When they are derived from more than two variables, the multidimensional meaning would be more difficult to interpret. Interpreting more than two variables is a goal of future work.

I will focus on two methods for interpreting the combined effect of correlated variables: omitting a variable and explicitly combining the correlated variables using principal component analysis. For clarity, my example model has only two substantially correlated variables with a theoretically meaningful joint effect on the

[2] Of course, my method only works when the variables have a reasonable level of correlation, enough to say that they move together in a meaningful way. If they were not, then the principal component direction would tell us very little about either original variable.

model. Principal component analysis as usually developed involves all independent variables in the model (Montgomery et al., 2012); my approach will apply principal component analysis only to the two substantially correlated variables.

3 CETERIS PARIBUS?

The interpretation of the coefficients in any multivariate regression model requires ceteris paribus; all else is equal. The assumption that all else is equal is used to understand the effect of each variable independent of any other variable. For many datasets, this is a very useful assumption. The utility of this assumption breaks down when two variables are highly correlated with each other. It can still be used, but only with careful consideration of the distribution of each correlated variable conditional on the other.

The ideal scenario for interpreting coefficients in a regression model is one in which each variable is uncorrelated (orthogonal) to the others. As an example, the ideal is to understand the power of money independent of any other possible confounding factor, such as knowledge of the political system. This is precisely what the coefficient on district income shows in a model that predicts legislative success. It shows how legislative success changes as district income changes, independent of the confounding impact of education. Yet the data generating process may not allow something like economic power to have much variation independent of the knowledge and network provided by educational attainment. Indeed, both may be caused by a common latent variable such as parental social standing, drive, or ability. In this situation, there may be very few districts with many highly educated constituents yet few high income constituents. That is, all else is not equal. Interpreting the effect of district income, ceteris paribus, may largely end up explaining the behavior of representatives from very few districts, or extrapolate beyond the observed data entirely. When the data generating process requires that one variable move in tandem with the other, or when it is not meaningful to extrapolate effects outside of the observed sample, it may be more meaningful to interpret the joint effect two correlated variables.

In this section, I construct a simulated dataset of 10,000 observations that demonstrates this principle. I then apply two techniques that allow the joint impact of both variables to be interpreted for both magnitude of effect and statistical significance. The first and cleanest technique applies a minor variation of principal component analysis. The other technique reinterprets the coefficients on retained variables in omitted variable models. Traditionally, the change in a coefficient when another variable is omitted is called omitted variable bias. I show that the coefficient in the omitted variable model can, and should, be interpreted as the joint effect of the retained and omitted variable. The simulated data will highlight how to interpret both the independent and joint impact of the correlated variables on the dependent variable. In later examples, I apply these methods to real datasets.

I define a multivariate regression model as follows. Assume that the true data generating process for the dependent variable Y is

$$\begin{aligned} Y &= \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \mu \\ &= X_2 + X_3 + X_4 + \mu, \qquad \mu \sim N(0, 20). \end{aligned} \tag{2}$$

Equation 3

$$\hat{Y} = \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3 + \hat{\beta}_4 X_4. \tag{3}$$

is Equation 2 rewritten in terms of fitted values of Y.

Let X_1 be a latent variable that causes both X_2 and X_3. I define the primary explanatory variables, X_2 and X_3 in terms of X_1:

$$X_1 \sim N(0,1)$$

$$X_2 = X_1 + 3 + \epsilon_0, \qquad \epsilon_0 \sim N(0, 0.2) \tag{4}$$

$$X_3 = X_1 + 2 + \epsilon_1, \qquad \epsilon_1 \sim N(0, 0.3) \tag{5}$$

Note that X_3 can be rewritten in terms of X_2,

$$\begin{aligned} X_3 &= \beta_{23} X_2 + c + \epsilon_1 - \epsilon_0 \\ &= X_2 - 1 + \epsilon_3, \qquad \epsilon_3 = \epsilon_1 - \epsilon_0. \end{aligned} \tag{6}$$

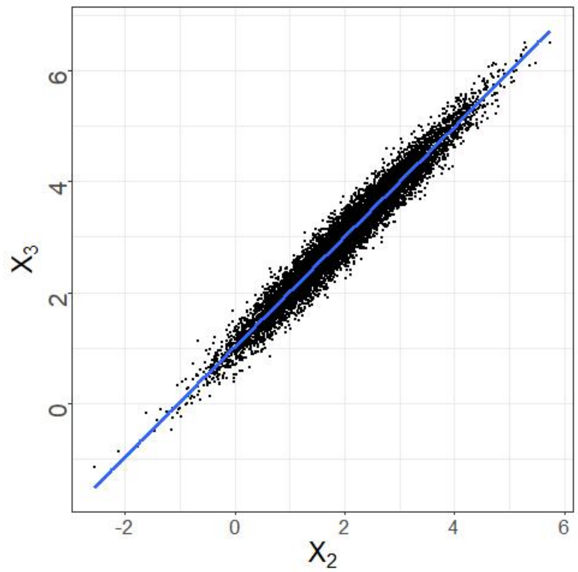
Let $X_4 \sim N(0,1)$ stand in for all controls in the model. Let X_2 and X_3 be the primary explanatory variables, defined in Equations 4 and 5. The true slope of the effect of X_2 is $\beta_2 = 1$, and the true effect of X_3 is represented by $\beta_3 = 1$.

The variables X_2 and X_3 are highly correlated by design, as seen in Figure 2. The regression results show an estimated slope for X_2 slightly under 1, which is the true value of β_{23}. The estimated regression line is plotted in blue.

The expected value of $\hat{X}_3$ as X_2 changes by one unit is defined to the slope, β_{23}. In this case the strong correlation between the two variables is a boon. The expected value of $\hat{X}_3$ can be estimated with a large degree of precision. Yet it is the strength of this correlation that induces tricky interpretations of the regression when both X_2 and X_3 are included as explanatory variables in a regression model.

Figure 3 shows what this means about how $\hat{Y}$ is expected to change as X_2 changes, ceteris paribus. In order to let all else be equal, we define $X_3 = 1$ and

Figure 2: Simulated Data, Scatterplot of X_2 and X_3 with Regression Line

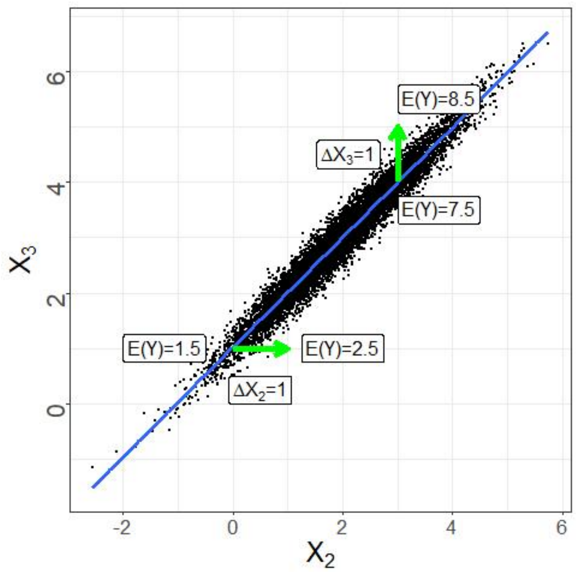


$X_4 = 0.5$ to be fixed values, then let $\Delta X_2 = 1$. When X_2 changes by one unit, $\Delta E(\hat{Y}) = 1$, as depicted in the lower half of Figure 3. The top half of the figure shows the same effect for ΔX_3 based on β_3. This time X_2 is held constant at the value 3 and $X_4 = 0.5$ as before. The expected change in $\hat{Y}$ when $\Delta X_3 = 1$ is also one, ceteris paribus. Yet note that both of these changes end up predicting points that are not in the observed sample. They interpretation is being made for data that may not exist in the real world.

The notation for partial derivatives and directional derivatives will be used in the section on principal components, so I introduce it here. $\hat{\beta}_2$ and $\hat{\beta}_3$ can be represented by both the partial derivative of Y and a special case of the directional derivative of Y. The partial derivatives of Y with respect to X_2 and X_3, respectively, are represented by

$$\hat{\beta}_2 = \frac{\partial \hat{Y}}{\partial X_2}, \hat{\beta}_3 = \frac{\partial \hat{Y}}{\partial X_3}. \tag{7}$$

Figure 3: Simulated Data with Scatterplot, Regression, and Expected $\Delta\hat{Y}$



Both partials assume that ceteris paribus, that all other variables are held constant. This can be explicitly described by the directional derivatives in the direction of the X_2 and X_3 axes. Consider the direction of the X_2 axis, represented by the unit vector $u = \begin{bmatrix} X_2 \\ X_3 \\ X_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$. Then $\hat{\beta}_2$ is the directional derivative, D_u, of $\hat{Y}$ in the direction $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$:

$$
\begin{aligned}
\hat{\beta}_2 &= D_u \hat{Y} \\
&= \begin{bmatrix} \frac{\partial \hat{Y}}{\partial X_2} \frac{\partial \hat{Y}}{\partial X_3} \frac{\partial \hat{Y}}{\partial X_4} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \frac{\partial \hat{Y}}{\partial X_2}, \qquad (8) \\
\hat{\beta}_3 &= D_u \hat{Y} \\
&= \begin{bmatrix} \frac{\partial \hat{Y}}{\partial X_2} \frac{\partial \hat{Y}}{\partial X_3} \frac{\partial \hat{Y}}{\partial X_4} \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \frac{\partial \hat{Y}}{\partial X_3}. \qquad (9)
\end{aligned}
$$

This section has demonstrated two main problems with the assumption ceteris paribus when two explanatory variables are highly correlated. The first is that the estimated effect is difficult to estimate precisely. Let X_2 change from zero to one as depicted in Figure 3. There are relatively few data points in this direction. This will result in inefficiencies in the estimation of β_2 because the number of useful observations are few. Another concern with using ceteris paribus is that the change predicted by the estimated coefficients only explain a small portion of the data. By moving approximately along the regression line from $(X_2, X_3) = (0, 1)$ changes can be predicted for the bulk of the data.

Given the correlation between X_2 and X_3, a more meaningful change in the data may be represented by changes that occur along the central axis of the observed data. This central axis is approximated by the regression line. A more exact description of this central axis is derived through principal component analysis. This central axis captures the expected change in $\hat{Y}$ when X_2 moves from $X_2 = 0$ to

$X_2 = 4$ by incorporating the inherent change in X_3 based on the observations in the data. Principal components also allow the standard error to be calculated using the tools available in any regression estimation package, as well as the statistical significance of the coefficient and the fit of the overall model.

4 PRINCIPAL COMPONENT ANALYSIS and JOINT EFFECTS

Principal component analysis and related methods seek to uncover underlying dimensionality among highly correlated variables. This has been widely used in political science to extrapolate ideological scores for politicians based on voting behavior and donation patterns (Poole & Rosenthal, 1997; Martin & Quinn, 2002; Bonica, 2014). The standard use of principle component analysis in linear regressions involves transforming all independent variables to discover the latent dimensionality (Montgomery et al., 2012). I develop an application of principal component analysis to test the statistical significance and interpret the substantive meaning of the joint effect of two correlated variables without changing any other variables in the model.

For expository purposes, I demonstrate this concept using the same simulated data developed in the previous section. Subsequent sections will apply this method to an analysis of the impact of education and income, two highly correlated explanatory variables, on a variety of political outcomes.

This section will divide the explanation of principal components into three: the first focuses on the transformation of X to create a coordinate system where the primary axis travels along the heart of the observed data. This axis will be similar, but not identical, to the regression line shown in Figure 2. The second explanation demonstrates that the estimated coefficients for this transformed data are a linear combination of the estimated coefficients from the original data. The last explanation will focus on how to apply these transformations to the simulated data.

Let $X_{23} = (X_2, X_3)$. Now define $Z_{23} = (X_2 - \bar{X}_2, X_3 - \bar{X}_3)$, which centers each column. The following linear transformation defines the rotation that produces X_{PCA}

$$\begin{aligned} X_{PCA} &= (X_{PC1}, X_{PC2}) && (10) \\ &= Z_{23}V && (11) \\ &= (X_2, X_3)V - (\bar{X}_2, \bar{X}_3)V && (12) \end{aligned}$$

where

$$V = \begin{bmatrix} a & -b \\ b & a \end{bmatrix} \tag{13}$$

and $a^2 + b^2 = 1$. Note that $V^{-1} = V^T$ (Montgomery et al., 2012). The values for a and b are chosen in a way that ensures the new matrix X_{PCA} has the following two properties:

- Each column in X_{PCA} is orthogonal to the other. This orthogonality is imposed by any values of a and b in V that satisfy $a^2 + b^2 = 1$.

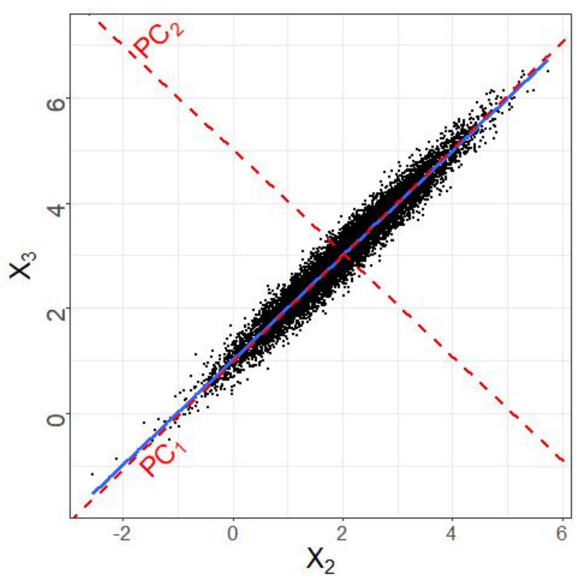
- The variance in the first column, X_{PC1} captures the greatest possible variance in the data. The second column, X_{PC2}, maximizes the variance possible to capture, subject to the constraint that it is orthogonal to the first column.

This maximization problem determines the values of a and b (Jolliffe, 2011).

Note that this can be generalized to any $n \times p$ matrix X. That is, a matrix V can be found that performs a linear transformation that rotates X to create X_{PCA} in a way that requires that each column of X_{PCA} is orthogonal to the other, and each column maximizes the possible variance subject to the constraint that it is orthogonal to all previous columns in X_{PCA}.

This transformation produces a new coordinate system for the observations in X_{23} pictured in Figure 4. The axes for this new matrix X_{PCA} can be plotted in the original X_2, X_3 plane, as seen in Figure 4. The values of a and b and the equations that define this new coordinate system will be described in more detail later in this section.

Figure 4: Simulated Data, Scatterplot with Lines for PC_1 and PC_2



This transformation into a new coordinate system for the data in X_{23} is associated with a clean and easily interpretable linear transformation of β in linear regressions. The general linear relationship in a least squares problem is written as

$$Y \cong X\beta + \beta_0, \tag{14}$$

where $X = (X_2, X_3, X_4)$ and $\beta = \begin{bmatrix} \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix}$. For simplicity of presentation, in the subsequent derivations I assume that X and Y have been centered, with the mean removed from all columns. The solution also applies when the mean is added back to X and Y, as the estimated coefficients of centered variables are the same as the estimated coefficients for uncentered variables (Montgomery et al., 2012) except for constant terms. This assumption implies that $\beta_0 = 0$ in the solution to Equation 2, and we can consider

$$Y \cong X\beta. \tag{15}$$

The regression solution to this relationship is

$$\hat{\beta} = (X^T X)^{-1} X^T Y. \tag{16}$$

I want to relate X_{23}, X_{PCA}, and V in the general regression equation. To do this, I define

$$V_{all} = \begin{bmatrix} a & -b & 0 \\ b & a & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{17}$$

and

$$
\begin{aligned}
X_{PCX} &= (X_{PC1}, X_{PC2}, X_4) \\
&= XV_{all}.
\end{aligned} \tag{18}
$$

More control variables can easily be accommodated by adding columns of the identity matrix to V_{all}.

Consider the solution to the least squares regression problem

$$
Y \cong X_{PCX}\beta_{PCX}. \tag{19}
$$

I will show that the regression solution to the problem in Equation 19 is $\hat{\beta}_{PCX} = V_{all}^T\hat{\beta}$:

$$
\begin{aligned}
\hat{\beta}_{PCX} &= (X_{PCX}^T X_{PCX})^{-1} X_{PCX}^T Y \\
&= ((XV_{all})^T XV_{all})^{-1}(XV_{all})^T Y \\
&= (V_{all}^T X^T XV_{all})^{-1} V_{all}^T X^T Y \\
&= V_{all}^{-1}(X^T X)^{-1} V_{all} V_{all}^T X^T Y \\
&= V_{all}^T (X^T X)^{-1} X^T Y \\
&= V_{all}^T \hat{\beta}.
\end{aligned} \tag{20}
$$

The linear combination of β_2 and β_3 that produce β_{PC1} and β_{PC2} are shown below:

$$\begin{aligned}
Y &\cong (X_{PC1}, X_{PC2}, X_4) \begin{bmatrix} \beta_{PC1} \\ \beta_{PC2} \\ \beta_4 \end{bmatrix} \\
&\cong \beta_{PC1}X_{PC1} + \beta_{PC2}X_{PC2} + \beta_4 X_4 \\
&\cong (a\beta_2 + b\beta_3)X_{PC1} + (-b\beta_2 + a\beta_3)X_{PC2} + \beta_4 X_4. \qquad (21)
\end{aligned}$$

This transformation leaves the column X_4 in X_{PCX} and X identical, and leaves β_4 unchanged in β_{PCX} and β. Thus, without loss of generality I focus on demonstrating the form of the 2×2 matrix X_{PCA} and the 2×1 vector $\beta_{PCA} = \left[\begin{smallmatrix} \beta_{PC1} \\ \beta_{PC2} \end{smallmatrix}\right]$.

I now focus on the principal component transformation as applied to the simulated data I created in the previous section. The first step is to find the values of V. I use the command prcomp() in the package "stats" in R to extract the following weights for the rotation matrix V

$$V = \begin{array}{c} \\ X_2 \\ X_3 \end{array} \!\!\begin{array}{c} \begin{array}{cc} PC_1 & PC_2 \end{array} \\ \begin{bmatrix} 0.7129 & -0.7013 \\ 0.7013 & 0.7129 \end{bmatrix} \end{array}. \qquad (22)$$

The two principal components are calculated from the original X_{23} matrix using Equation 18. I add in an additional row to calculate the translation required for X_{PCA} to be centered at the mean of X_{23}. This will allow me to fully define the principal component axes plotted in the X_2, X_3 plane shown in Figure 4.

$$X_{PCA} = (1, X_2, X_3)\begin{bmatrix} -a\bar{X}_2 - b\bar{X}_3 & b\bar{X}_2 - a\bar{X}_3 \\ 0.7129 & -0.7013 \\ 0.7013 & 0.7129 \end{bmatrix} \tag{23}$$

$$= \begin{bmatrix} 0.7129X_2 + 0.7013X_3 - (a\bar{X}_2 + b\bar{X}_3) \\ -0.7013X_2 + 0.7129X_3 - (b\bar{X}_2 - a\bar{X}_3) \end{bmatrix}^T \tag{24}$$

$$= \begin{bmatrix} X_{PC1} \\ X_{PC2} \end{bmatrix}^T \tag{25}$$

In short, each principal component is calculated by the following linear combinations:

$$PC_1 = aX_2 + bX_3 - (a\bar{X}_2 + b\bar{X}_3) \tag{26}$$

$$= 0.7129X_2 + 0.7013X_3 - 3.5297$$

$$PC_2 = -bX_2 + aX_3 - (-b\bar{X}_2 + a\bar{X}_3) \tag{27}$$

$$= -0.7013X_2 + 0.7129X_3 + 0.7361$$

The axes for this new coordinate system can be plotted with respect to the original coordinate plane (X_2, X_3). The equation for these axes are defined by $PC_1 = 0$ in Equation 28 and $PC_2 = 0$ in Equation 29, and are plotted in red in Figure 4.

Axis for the first principal component

$$X_3 = \frac{b}{a}X_2 + (\bar{X}_3 - \frac{b}{a}\bar{X}_2) \tag{28}$$
$$= 0.9837 * X_2 - 1.0325$$

Axis for the second principal component

$$X_3 = -\frac{a}{b}X_2 + (\frac{a}{b}\bar{X}_2 + \bar{X}_3) \tag{29}$$
$$X_3 = -1.0165 * X_2 + 5.0331$$

The transformations to create X_{PCA} from each observation in X_{23} in the simulated data, as well as the equations for the new principal component axes with respect to the original (X_2, X_3) plane have been defined. Two steps remain: calculate the new values in β_{PCA} based on the original β_{23} estimates, and use these calculations to demonstrate the estimated changes in Y as X_2 and X_3 change in the direction of $PC_1 = \left[\begin{smallmatrix} X_2 \\ X_3 \end{smallmatrix}\right] = \left[\begin{smallmatrix} a \\ b \end{smallmatrix}\right]$.

The equations for the new $\hat{\beta}_{PCA}$, defined by Equation 21, are described below.

$$
\begin{aligned}
\hat{\beta}_{PCA} &= \begin{bmatrix} \hat{\beta}_{PC1} \\ \hat{\beta}_{PC2} \end{bmatrix} \\
&= V^T \hat{\beta}_{23} \\
&= \begin{bmatrix} a & b \\ -b & a \end{bmatrix} \begin{bmatrix} \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} \\
&= \begin{bmatrix} a\hat{\beta}_2 + b\hat{\beta}_3 \\ -b\hat{\beta}_2 + a\hat{\beta}_3 \end{bmatrix} \\
&= \begin{bmatrix} 0.7129\hat{\beta}_2 + 0.7013\hat{\beta}_3 \\ -0.7013\hat{\beta}_2 + 0.7129\hat{\beta}_3 \end{bmatrix}
\end{aligned} \tag{30}
$$

In short, $\hat{\beta}_{PCA}$ is a linear combination of the original $\hat{\beta}$ values. The principal component, $\hat{\beta}_{PC1} = a\hat{\beta}_2 + b\hat{\beta}_3$, is a particularly interesting combination as it captures the way most of the observations behave in the data. The stronger the correlation between X_2 and X_3, the more meaningful it will be to examine changes along that PC_1 axis.

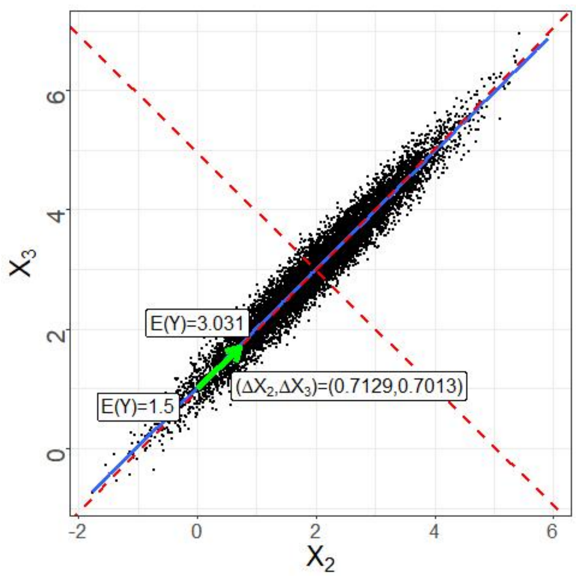
The substantive meaning of the combination in $\hat{\beta}_{PC1}$ is the expected change in $\hat{Y}$ as we move through the data in the direction $\left[\begin{smallmatrix} a \\ b \end{smallmatrix}\right]$. This is the direction that has the slope $\frac{b}{a}$ in Figure 4, which is also the direction that moves in same way as the bulk of the observed data. Namely, it is the direction captured by the slope of the first principal component line, which has the slope $\frac{b}{a}$.

Figure 5 shows the estimated change in Y as X_2 and X_3 both increase by one unit. Here we can see that

$$E(\Delta Y|(\Delta X_2, \Delta X_3) = (a, b)) = \hat{\beta}_{PC1}, \tag{31}$$

where $\hat{\beta}_{PC1}$ is the calculated coefficient, 1.531, for the first principal component vector from Table 1. This can also be understood in terms of directional derivatives.

Figure 5: Simulated Data with Scatterplot, Regression, and Expected ΔY along Principal Component



Based on Equation 20, $\hat{\beta}_{PCA} = V^T\hat{\beta}_{23}$. Recall that the partial derivatives of each variable take the following form in the estimated model:

$$\hat{\beta}_i = \frac{\partial \hat{Y}}{\partial X_i} \tag{32}$$

Focusing on $\hat{\beta_{PC1}}$,

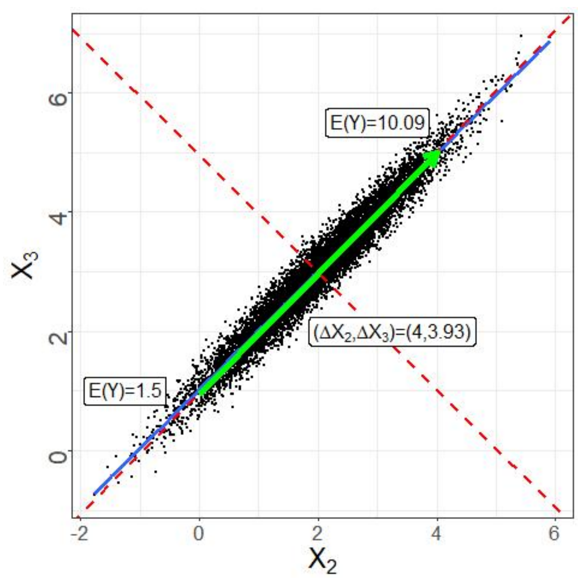
$$\hat{\beta}_{PC1} = a\hat{\beta}_2 + b\hat{\beta}_3 \tag{33}$$

This can be rewritten as a linear combination of partial derivatives.

$$\hat{\beta}_{PC1} = a\frac{\partial \hat{Y}}{\partial X_2} + b\frac{\partial \hat{Y}}{\partial X_3} \tag{34}$$

This is a directional derivative, and simply represents the rate of change of $\hat{Y}$ in the direction $\left[\begin{smallmatrix} X_2 \\ X_3 \end{smallmatrix}\right] = \left[\begin{smallmatrix} a \\ b \end{smallmatrix}\right]$. That is, $\hat{\beta}_{PC1}$ allows us to estimate the magnitude of the effect on the change in Y of changing X_2 and X_3 as we move along the first principal component direction.

Figure 6: Simulated Data with Scatterplot, Regression, and Expected ΔY along Length of Principal Component



One benefit of the principal component is that the estimated effect of the longest spread of the data can be estimated, as approximated in Figure 6. When interpreting the independent effect (partial derivatives) of either X_2 or X_3, one needs to be aware of the limitations inherent in the reduced spread of the data in the conditional distribution of one variable on the other. Comparing the expected value of Y at the low end of the joint distribution of X_2 and X_3 against the expected value at the high end, a change of 8.5 in Y is predicted.

The principal component offers multiple obvious benefits. It allows the researcher to compare estimated effects for observations that exist in the data by examining the net effect of multiple variables when the independent (or marginal) effect of each variable is minimal given the spread of the data. In this model, it is entirely plausible that many observations would expect to be associated with data points along the green arrow, of length approximately 6 in Figure 6. Yet while this is possible in the direction of the principal component, it is not possible in the direction of X_2 holding all else constant. In that direction, the change in X_2 cannot go much beyond 1 unit.

Additionally, precision of the estimate for PC_1 is likely to be substantially higher than the precision of the estimates for $\hat{\beta}_2$ or $\hat{\beta}_3$. The spread of one correlated variable is low conditional on the other variables. This effectively leads to a much smaller number of observations for the estimation, so the precision of the estimate will be low; there are many more observations in the direction of the first principal component which should produce higher precision estimates. This is the subject of the subsequent sections that discuss when multicollinearity bites.

5 REINTERPRETING OMITTED VARIABLE BIAS

It is well known that when two or more variables are highly correlated and one or more are omitted from the model, the coefficients on the retained variables can meaningfully change. They are no longer unbiased estimates of the original full model coefficients. This can range from mildly altering the size of the coefficients on the remaining variables to changing the sign and statistical significance. This has typically been interpreted as a bias caused by an omitted variable. When this leads to misinterpreted results, confounding is said to occur. The traditional cure has been to avoid omitting variables, or to recognize that the retained variables are necessarily subject to all types of confounding factors that cannot be or have not been included in the model (Clarke, 2005).

When two variables are known to be highly correlated and one variable is omitted, I propose an alternative view of the regression coefficient on the remaining variable. Critically, I reinterpret the coefficients on the remaining variables as the joint effect of each retained correlated variable with the omitted variable. Omitted variable bias only occurs when these coefficients are not reinterpreted. This may be unavoidable if the omitted variables are completely unknown. But when the omitted variable is known, and especially when it has been measured, the coefficients on the retained variables can be meaningfully interpreted as the joint impact of the omitted and retained variable.

To show how the coefficients can be reinterpreted, I examine the simplest case where there are only two explanatory variables in the model. Greene (2003) discusses the more general case. Let the true relationship be represented by Equation 35, where the β terms are the coefficients of the explanatory variables X_1 and X_2. Y is the dependent variable, c is the constant and μ is the error term.

$$Y = \beta_1 X_1 + \beta_2 X_2 + c + \mu \tag{35}$$

If X_1 and X_2 are collinear, then X_1 can be linearly predicted by X_2 with a substantial degree of accuracy. This is be represented by Equation 36, where β_{1a} is the calculated regression coefficient:

$$X_2 = \beta_{1a} X_1 + d + \epsilon \tag{36}$$

where ϵ is small.

The effect of dropping X_2 from Equation 35 can be demonstrated by substituting Equation 36 into Equation 35. This substitution is presented in Equation 37.

$$Y = (\beta_1 + \beta_{1a}\beta_2)X_1 + (\beta_2 d + c) + (\beta_2 \epsilon + \mu) \tag{37}$$

Let the new coefficient on X_1 be

$$\beta_1^{\cancel{X2}} = \beta_1 + \beta_{1a}\beta_2. \tag{38}$$

The approach used here highlights that $\beta_1^{\cancel{X2}}$ is not the direct effect of the independent impact of X_1 on Y. Instead, it is the sum of the direct effect, β_1 plus the indirect effect of X_2, β_2, channeled through the correlation between X_1 and X_2, β_{1a}. Note that the independent effect of the omitted variable X_2 is captured in both the new coefficient on X_1 and in the error term $\beta_2\epsilon + \mu$.

This also makes it clear how confounding can happen. If the magnitude of $\beta_{1a}\beta_2$ is large relative to β_1, then the influence of β_2 can move the value of $\beta_1^{\cancel{X2}}$ far from the value of β_1. If the sign of β_2 or β_{1a} opposes the sign of β_1, then the sign of $\beta_1^{\cancel{X2}}$ can be different from the sign of β_1. This is a substantial problem if $\beta_1^{\cancel{X2}}$ is assumed to be β_1, as it will lead to the incorrect inference that the independent effect of X_1 on Y is positive when the true value of β_1 is negative, or vice versa. This misinterpretation is also referred to as omitted variable bias (Clarke, 2005). Yet if considered properly as distinct from β_1 and if the joint impact of the multicollinear variables is theoretically meaningful, β_1^{X2} is interpretable for both statistical significance and substantive meaning.

This coefficient is interpreted as the sum of the direct effect of X_1 on Y and the indirect effect of X_2 on Y through the relationship of X_1 and X_2. In less technical terms, it is the joint effect of both X_1 and X_2 on Y. In a more complex case, it would be the joint effect of all dropped variables on the remaining variables in the model.

To restate, when a variable, call it X_2, is omitted from the model but is collinear with a retained variable X_1, the coefficient of the retained variable is no longer reflective of its independent effect on Y. If the assumption is that the coefficient reflects this independent effect, then it is fair to say that the coefficient is biased away from the independent effect. Yet the coefficient is not biased away from the joint effect of X_1 and X_2 on Y; it is an unbiased estimate of Equation 38, $\beta_1 + \beta_{1a}\beta_2$ (Clarke, 2005).

The decision of which variable to keep can be made based on which variable is more theoretically relevant. Yet it is critical to interpret the coefficient on the retained variable as the combined effect of both variables. Specifically, it the sum of the direct effect of the retained variable and the indirect effect of the omitted variable.

A word of caution is required for this method. The models discussed in this section assumed that the true data generating process only involved two explanatory variables. Yet almost all estimated models contain many more independent variables. For models with more variables, omitting a variable will impact all retained variables. The degree to which they are impacted depends on their correlation with the omitted variable and the magnitude of the true effect of that omitted variable. If the coefficients on the controls are not substantively important, or if the omission does not meaningfully change the interpretation of those coefficients, then the reinterpretation I argue for here is largely applicable. If either is not true, then care must be taken to properly interpret the remaining coefficients as the joint impact of the retained variable with the omitted variable.

This method for interpreting the joint effect of two variables is not as reliable as the method provided by principal component analysis. Unlike PCA, any other retained variables must also be reinterpreted in light of the omitted variable. Additionally, in PCA the direction of the estimated effect of joint was explicitly $\left[\begin{smallmatrix} X_2 \\ X_3 \end{smallmatrix}\right] = \left[\begin{smallmatrix} a \\ b \end{smallmatrix}\right]$. The interpretation of the direction of the effect of $\beta_1^{\cancel{X2}}$ is less straightforward,[3] although it will be similar to the effect of the first principal component if X_1 and X_2 are highly correlated.

The benefit to omitted variables is that it is simple to implement if care is taken with how the results are interpreted. It can be used as a robustness check on the results, or as part of the exploratory phase of any data analysis project.[4] Omitting variables from a model is one way to detect if a primary explanatory variable is sufficiently correlated with another independent variable to either (a) affect the coefficient on the primary explanatory if the other correlated variable is omitted or (b) examine jointly to understand the net effect of both variables. If the omission of a variable can substantially impact the coefficients of the retained variables, this is worth highlighting and examining in its own right.

Two other methods for creating a combination of multicollinear variables were

[3]Future research may clarify the direction of the effect.

[4]The researcher should, as always, clearly divide hypothesis testing from data exploration. Both are valuable, but when they are mixed it is easy to unintentionally p-hack.

discussed in Section 2: sequential residual regression, clustering, and structural equation modeling (Graham, 2003). At their core, although they are frequently interpreted as estimating the independent effect of one variable, they are all variations on estimating coefficients that reflect the joint impact of all or some of the multicollinear variables.

This section and the previous section show the utility of principal components and omitted variable model to calculate the joint impact of two variables. From this joint impact, hypotheses can be tested that are both more specific and more relevant to the observed data and data generating process. Another benefit to these methods occurs in the special case where multicollinearity obscures the independent impact of each multicollinear variable.

6 MULTICOLLINEARITY and OBSCURED RESULTS

Multicollinearity is well known to distort the effects of variables when two or more collinear variables are included in a model.[5] Less well known are the limitations of that distortion. Multicollinearity cannot create false statistically significant results. However, it can obscure the independent effect of a variable by causing inflated standard errors. The methods I have proposed to test the joint impact of two correlated variables is useful to apply to the case where multicollinearity obscures the independent effects of those variables.

One misconception is that the $\hat{\beta}$ estimator, i.e. the regression coefficients, can

[5]The manual for the statistical package NCSS says: "Multicollinearity can create inaccurate estimates of the regression coefficients, inflate the standard errors of the regression coefficients, deflate the partial t-tests for the regression coefficients, give false, non-significant, p-values, and degrade the predictability of the model (and that's just for starters)." (*Ridge Regression*, n.d.)

be biased by collinearity. This is false. The proof that $\hat{\beta}$ is unbiased does not require the dependent variables be uncorrelated. It will require that the error terms have a mean of 0, are identically distributed, and are uncorrelated, but it will not put any requirements on the collinearity of the matrix of independent variables, X. That is, the expected value of $\hat{\beta}$ is identical to the true value of β regardless of collinearity.

The confusion about collinearity's effect on regression estimates comes from a separate yet potentially confusing proof: the least squares distance between the true values of β and the $\hat{\beta}$ values is inflated when collinearity is present. It is an artifact of calculating distance, which involves squaring the difference between the true value beta and the estimated beta hat coefficients. This can be interpreted as indicating the un-squared $\hat{\beta}$ is too large, but it is not. Multicollinearity does not cause $\hat{\beta}$ estimates to be biased.

The primary influence of multicollinearity is on the confidence intervals. This affects the precision of the estimates, not the accuracy.

> In effect, severe multicollinearity inflates the variances of the regression coefficients, and this increases the probability that one or more regression coefficients will have the wrong sign (Montgomery et al., 2012, p. 121).

Of course, the prediction about a coefficient having the wrong sign is only valid when the variances are inflated. That is, the coefficient may be both not statistically significant and have the wrong sign. Collinearity alone will not produce a statistically significant coefficient with the wrong sign. If the results are statistically significant, then it is at least as correct as any effect seen in the absence of multicollinearity. As Belsley succinctly put it:

> Thus, if an investigator is only interested in whether a given coefficient is significantly positive and is able, even in the presence of collinearity, to accept that hypothesis on the basis of the relevant t-test, then collinearity has caused no problem. Of course, the resulting forecasts or estimates may have wider confidence intervals than would be needed to satisfy a more ambitious

> researcher, but for the limited purpose of the test of significance initially proposed, collinearity has caused no practical harm. These cases serve to exemplify the pleasantly pragmatic philosophy that collinearity doesn't hurt so long as it doesn't bite (Belsley, 1991, p. 73).

The best case scenario for a researcher is the one in which multicollinearity does not bite. Although the statistical significance of the estimated coefficients may be inflated, the accuracy of $\hat{\beta}$ is not impacted and the null hypothesis is rejected. The researcher can estimate the substantive effect of the variables under the normal guidelines of hypothesis testing, i.e. ceteris paribus. Yet as I have shown in prior sections, all else may not be equal. Understanding the joint impact of the collinear variables can still be substantively important.

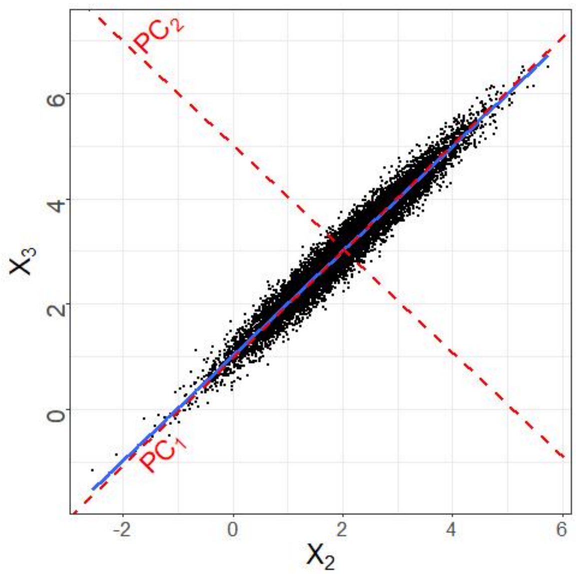
7 SIMULATED MULTICOLLINEARITY

This section will examine the interpretation of a simulated case where multicollinearity bites (Case 3) against a case where multicollinearity exists but does not bite (Case 1). In each case I show regression output for Models 1-4. I then run 1,000 simulation of the model in Equation 2 and plot the estimated coefficients $\hat{\beta}_2$ and $\hat{\beta}_3$ to show the effect of multicollinearity on the distribution of the estimated coefficients for those multicollinear variables.[6]

I use the same simulated data developed in the section "Ceteris Paribus?" for the first example. For the second simulated example, I use the same (X_2, X_3, X_4) variables but create a new dependent variable, Y_2. The scatter plot output of the original two explanatory variables, X_2 and X_3, along with the principal component axis, is replicated in Figure 7 As demonstrated in the section on principal compo-

[6]Since X_2 and X_3 are defined to be collinear, I do not perform the traditional diagnostics for collinearity in this section. The first applied section, examining legislative success based on district income and education, demonstrates a number of diagnostics for detecting collinearity.

Figure 7: Scatter plot of X_2 and X_3 with PCA Axis



nents, this graphic can be referenced to visualize the independent effect of X_2, X_3, and the joint effect of both along the PC_1 axis.

For each of the 1,000 trials in the simulation, I chose Y according to the data generating process in Equation 2 and estimated β using a regression. I focus on $\hat{\beta}_2$ and $\hat{\beta}_3$ from each trial in the simulation. I plot the set of values $(\hat{\beta}_2, \hat{\beta}_3)$. This provides a picture of the set of potential $\hat{\beta}_2$ and $\hat{\beta}_2$ values as Y varies according to the data generating process in Equation 2. Figure 7 demonstrates an example of this plot.

7.1 Multicollinearity Bites

In the first example, the dependent variable is created by the following the same data data generating process as in prior sections, Model 1:

$$Y = X_2 + X_3 + X_4 + \mu. \tag{39}$$

As before, X_2 and X_3 are highly correlated with each other, but not with X_4. Here multicollinearity bites, as shown in Figure 8.

Figure 8: Simulated Coefficients for X_2 and X_3

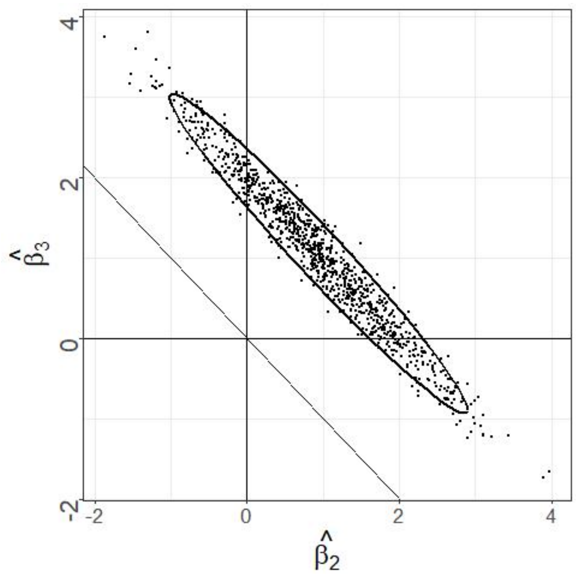


Figure 8 shows the results from the simulation of 1,000 coefficients for β_2 and β_3 from Model 1. An ellipse that captures 95% of these point estimates is superimposed on the scatter plot of point estimates.

The fact that multicollinearity bites is evident from the cigar shape of the distribution of the point estimates: long, thin, and crossing the $\hat{\beta}_2 = 0$ and $\hat{\beta}_3 = 0$ axes. This cigar shape demonstrates a fundamental fact of collinear data (Montgomery et al., 2012). Although the center of the ellipse and the mean of each estimate coefficient is unbiased: $(E(\hat{\beta}_2), E(\hat{\beta}_3)) \cong (1, 1)$. Yet $\hat{\beta}_1$ and $\hat{\beta}_2$ are correlated with each other. The more collinear the underlying data, the longer and thinner the cigar will be. The longer the cigar is, the more likely it is to extend over the $\hat{\beta}_2 = 0$ axis or the $\hat{\beta}_3 = 0$ axis. Twenty-seven percent of the point estimate pairs have one negative point estimate. In this simulation $\hat{\beta}_2$ was negative for 15.1% of the point estimates.

Thus, the simulation indicates that the p-value for $\hat{\beta}_2$ is likely to be in the vicinity of $p = 2 * .151 = .302$.[7] Thus, one cannot make any conclusions about the signs of $\hat{\beta}_2$ or $\hat{\beta}_3$.

The fact that multicollinearity bites is also evident in the standard regression model. Table 1 shows the regression results for one typical value of Y. There would be a similar, but not identical, table for each of the 1,000 Y values in the simulation. In this case, the coefficients $\hat{\beta}_2$ and $\hat{\beta}_3$ are not close to meeting the traditional threshold for statistical significance. Multicollinearity has produced variances that are so large that the signs of $\hat{\beta}_2$ and $\hat{\beta}_3$ are obscured.

Table 1. Collinearity Example: Multicollinearity Bites

	Model 1	Model 2	Model 3	Model 4
X2	1.447	2.134***		
	(0.898)	(0.197)		
X3	0.713		2.143***	
	(0.910)		(0.199)	
PC1				1.531***
				(0.141)
PC2				0.509
				(1.271)
X4	0.968***	0.970***	0.964***	0.968***
	(0.200)	(0.200)	(0.200)	(0.200)
Observations	10,000	10,000	10,000	10,000
Note: OLS			*p<0.05; **p<0.01; ***p<0.001	

Even though this is the worst possible case, where the signs of $\hat{\beta}_2$ and $\hat{\beta}_3$ are undetectable, we can still make conclusions about the substantive meaning and statistical significance of this data by using the principal components. Figure 9 shows the same point estimates transformed into the PCA coordinate system. The

[7] The value is doubled to make it equivalent to a two tailed test.

key point is that $\hat{\beta}_{PC1}$ is always positive. That is, the expected change in $\hat{Y}$ is positive in the direction of the PC_1 axis.

Figure 9: Simulated Coefficients for PC_1 and PC_2

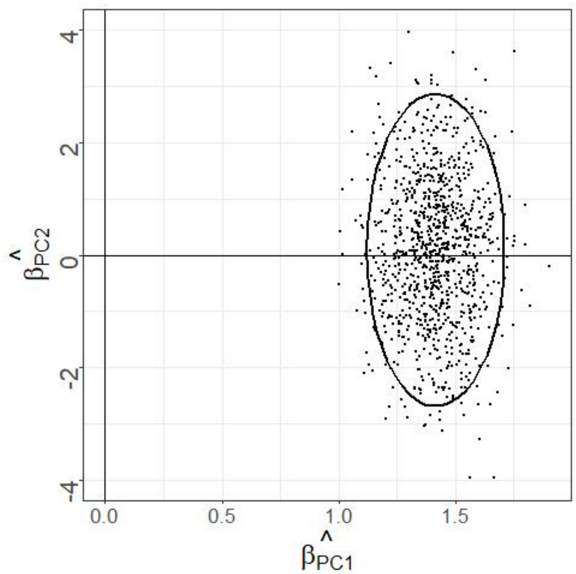


Table 1 shows the results for the principal component regression in Model 4. This provides additional information. We expect the 95% confidence interval for $\hat{\beta}_{PC1} = 1.531 \pm 1.96 * 0.141 = (1.25, 1.8)$. Also, we reject $H_0 : \beta_{PC1} = 0$ with high confidence. This has established the statistical significance. Substantively, it is the expected change in $\hat{Y}$ in the direction of PC_1, which was shown to be $u = \begin{bmatrix} X_2 \\ X_3 \\ X_4 \end{bmatrix} = \begin{bmatrix} 0.7129 \\ 0.7013 \\ 0 \end{bmatrix}$ in Equation 22. The coefficient $\hat{\beta}_{PC1}$ indicates how quickly Y changes as we move in the direction of the heart of the data.

As described in the section on omitted variables, Models 2 and 3 provide similar information about the joint effect of X_2 and X_3. The principal components have the advantage that all information originally in the full model (Model 1) is retained.

In this simulation, multicollinearity bit: neither $\hat{\beta}_2$ nor $\hat{\beta}_3$ were statistically significant. Yet using principal component analysis showed that the joint effect of

X_2 and X_3 is statistically significant and has substantive meaning.

7.2 Multicollinearity Does Not Bite

In the second example, the dependent variable is created using the following data generating process:

$$Y_2 = 4 * X_2 + 4 * X_3 + X_4 + \mu. \tag{40}$$

Table 2. Collinearity Example: Multicollinearity Does Not Bite

	Model 1	Model 2	Model 3	Model 4
X2	3.699***	7.571***		
	(0.896)	(0.196)		
X3	4.018***		7.673***	
	(0.908)		(0.199)	
PC1				5.455***
				(0.141)
PC2				−0.260
				(1.268)
X4	1.023***	1.035***	1.013***	1.023***
	(0.200)	(0.200)	(0.200)	(0.200)
Observations	10,000	10,000	10,000	10,000

Note: OLS *p<0.05; **p<0.01; ***p<0.001

All other variables are the same as before. In this case, Table 1 shows that multicollinearity is present but does not bite. The coefficients $\hat{\beta}_2$ and $\hat{\beta}_3$ in Model 1 of Table 2 are strongly statistically significant. Naturally, they are also jointly significant.

This demonstrates the case where the independent effect of each coefficient is visible despite multicollinearity. The precision of $\hat{\beta}_2$ and $\hat{\beta}_3$ are reduced, but the null hypothesis $H_0 : \beta_2 = 0$ or $H_0 : \beta_3 = 0$ is still overwhelmingly rejected.

8 MODEL INTERPRETATION

I present a guideline for interpreting models with non-orthogonal variables that provides information about the significance and substantive importance the independent contribution of each variable, as well as their joint contribution. This guideline highlights the importance of interpreting the joint impact of highly correlated variables.

For simplicity, I focus on the case where only two variables in the model have substantially collinearity. Let X_1 and X_2 be two highly correlated explanatory variables. For simplicity of presentation, I include one additional independent variable. All results apply equally well when there are many additional independent variables. Assume that the null hypothesis $\beta_1 = \beta_2 = 0$ is rejected, indicating that a combined effect of income and education may be statistically significant. I present guidelines for model interpretation in three cases, each drawn from one of the three possible combinations of statistically significant coefficients for X_1 and X_2 in the following model:

$$Y \sim \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \tag{41}$$

This guideline allows the magnitude of effect of both X_1 and X_2, either individually or combined, to be interpreted under traditional rules of statistical significance.

Case 1

Scenario The independent effects of both variables are statistically significant despite high collinearity. This is the best case scenario for a scholar interested in understanding the impact of education and income, or any other highly correlated variables.

Model Interpretation Both highly correlated variables can be interpreted, both individually and jointly. The following model will reveal the independent impact of each variable:

$$Y \sim \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3. \tag{42}$$

Yet this model may not reveal the full story. The net effect may be larger or smaller than the individual effects. Indeed, if X_1 and X_2 have countervailing effects then the net effect of their combination can be anything: null, positive, or negative. This would lead to confounded interpretations of the coefficients if one were omitted from the model. Since X_1 and X_2 are highly correlated, the net effect will be much more commonly observed in the data than the independent effects. Principal components can reveal the net effect. As discussed, this reflects the largest change in the observed data and therefore the heart of the meaningful changes in the real world. Let the principal component be X_{PC1}, the orthogonal component be X_{PC2}, and the model be:

$$Y \sim \beta_{PC1} X_{PC1} + \beta_{PC2} X_{PC2} + \beta_3 X_3. \tag{43}$$

Note that β_3 from both models will be identical because X and X_{PCA} contain the same information.[8] Then β_{PC1} reflects the net effect of X_1 and X_2 along the direction of the largest spread of the data. This allows the expected change in the dependent variable to be estimated from the difference between the most distinctive observations: those at the high

[8] Mathematically, the span of (X_{PC1}, X_{PC2}) is the same as the span of (X_1, X_2).

end of the first principal component, and those at the low end.

Note that it is fully possible for the joint effect of X_1 and X_2 to be statistically insignificant in this scenario if X_1 and X_2 are countervailing. If this happens, then the effect of X_1 has traditionally been interpreted as confounding $\hat{\beta}_2$. Using principal components reveals that this confounding is caused by X_1 and X_2 countering each other in a way that drives $\hat{\beta}_{PC1}$ very close to zero. Interpretation of the effect of each variable is limited in this scenario. One can conclude that X_1 and X_2 are individually significant, so increasing one while holding the other constant will have a statistically significant effect. Yet the substantive meaning can only be interpreted in so far as X_1 can change without X_2 also changing, or vice versa. There can be no meaningful conclusion about the difference between observations at the high end of both X_1 and X_2 as compared to observations at the low end. This method highlights meaningful information from this null result; the lack of a statistically significant conclusion is induced by the countervailing effects of the correlated variables.

<u>**Case 2**</u>

Scenario One of the two coefficients, β_1, is statistically significant and the other, β_2, is not.[9]

Model Interpretation The traditional approach to this scenario is to drop the statistically insignificant variable from the model.

$$Y \sim \beta_1^{\cancel{X2}} X_1 + \beta_3^{\cancel{X2}} X_3 \tag{44}$$

[9] A similar method applies if β_2 is statistically significant but not β_1.

Because statistical insignificance does not imply that β_2 is small,(Arceneaux & Huber, 2007) the new $\beta_1^{\cancel{X2}}$ is interpreted to include the combined effect even in the case where X_2 is insignificant. If β_2 is small, then the joint effect reflected in the reduced model $\beta_1^{\cancel{X2}}$ will be close to the value of β_1 in the full model. If β_2 is large, then the joint effect reflected in the reduced model will be closer to the coefficient on X_{PC1}. Using the principal component and full models, in Equations 42 and 43, would provide clarity to which it is.

Case 3

Scenario The independent impacts of both X_1 and X_2 are obscured because the model cannot reject the null hypothesis that either β_1 or β_2 is differentiable from zero.

Model Interpretation In this scenario, it is statistically impossible to differentiate the independent effect of X_1 from X_2. Yet it is still possible to derive meaningful interpretations, so long as the goal is to understand the joint impact of both variables. Three ways to do this, based on methods I have proposed in previous sections, are as follows:

- Use principal component analysis, or another similar technique, to create a new variable X_{PC1} that combines the impact of X_1 and X_2.

$$Y \sim \beta_{PC1} X_{PC1} + \beta_{PC2} X_{PC2} + \beta_3 X_3.$$

  This will typically be a linear combination of X_1 and X_2. As discussed in the section on principal component analysis, there are multiple benefits to using β_{PC1}. The biggest advantage is that it allows the estimated change in $\hat{Y}$ to be calculated based on the unconditional range of the distributions of X_1 and X_2.

- Use a model that omits X_1:

$$Y \sim \beta_2^{\cancel{X1}} X_2 + \beta_3^{\cancel{X1}} X_3 \tag{45}$$

  The key is that $\beta_2^{\cancel{X1}}$ represents the combined impact of X_1 and X_2, and should be interpreted as such.

- Use a model that omits X_2:

$$Y \sim \beta_1^{\cancel{X2}} X_1 + \beta_3^{\cancel{X2}} X_3 \tag{46}$$

  The interpretation is similar to dropping X_1.

The four models described are listed below:

$$Y \sim \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \tag{47}$$

$$Y \sim \beta_1^{\cancel{X2}} X_1 + \beta_3^{\cancel{X2}} X_3 \tag{48}$$

$$Y \sim \beta_2^{\cancel{X1}} X_2 + \beta_3^{\cancel{X1}} X_3 \tag{49}$$

$$Y \sim \beta_{PC1} X_{PC1} + \beta_{PC2} X_{PC2} + \beta_3 X_3. \tag{50}$$

I advocate comparing coefficients across multiple models, which can be difficult. One traditional approach is to standardize coefficients. Yet this can mislead interpretation of (1) the magnitudes of the coefficients, (2) whether the variables are fundamentally similar or are like comparing apples to oranges, and (3) comparisons between groups with differing spreads in the observations (King, 1986). I develop and present a visualization that address these points when I applied the methods described to real world data. It allows the reader to quickly examine both the magnitude of the effect of the coefficient as well as the statistical significance. The statistical significance is relevant to determine how multicollinearity impacted the various models, and therefore which models to are appropriate. Showing the magnitude of the effect allows the reader to examine the impact of omitted variable bias, as well as the substantive effect of each explanatory variable.

The following sections apply these guidelines and implement these graphics using the collinearity between education and income.

9 COLLINEARITY: EDUCATION and INCOME

I show evidence from four political outcomes that the independent impact of education is meaningful, and the joint impact of education is often more meaningful. Analyzing each on its own has been the standard, and has produced many useful results. Yet including one or both in a regression analysis without sufficient understanding of the ways their collinearity can impact the interpretation of the model can easily lead to misinterpreted effects. Specifically, the following misinterpretations are common:

- If education is omitted from the model but has an independent impact, then the coefficient for income must be interpreted as the joint effect of income and education.

- If income and education are retained in the model but are not statistically significant, they may be misinterpreted to have no effect on the dependent variable. Yet to truly determine this, their joint significance should be examined. If they are jointly significant, it can be substantively meaningful to interpret that joint effect.

- If income and education are retained in the model and are statistically significant, their joint impact may be more meaningful to interpret than the marginal effect of each in the model. I show examples where the net effect of income and education is negligible even though the independent effects are strongly significant. While the independent effects are meaningful, the joint effect reflects the bulk of the data. For example, most of the time a Republican from a high income district will appear to have the same chance of being elected as a Republican from a low income district. Yet conditional on a given level of district educational attainment, a Republican from a high income district will be more likely to be elected. The key is that there are not that many districts with substantially higher incomes than the level of education would predict. There are many districts that are simultaneously more educated and high income than others, and the effect of district education dramatically mitigates the effect of district income.

To show this I examine data on legislative success, partisan affiliation, and replicate results on country turnout rates (Burden & Wichowsky, 2014) and ideology (McCarty, Poole, & Rosenthal, 2006). I reveal insights by examining education and income with a better theoretical grounding on the conflicting requirements of multicollinearity and omitted variable bias.

Unlike income, race, and gender, education does not have a strong line of scholarly study. It is so closely tied to income, which has been rightly viewed as the predominant driver of political outcomes in the modern era, that it seems to have been dismissed as inconsequential. Not many studies include it in their models, and when they do it is rarely analyzed in any detail. (McCarty et al., 2006) show that

Republicans have lost many educated voters while they have gained high income voters over the past few decades. This could imply that they have ceased to be responsive to the educated as well, and this is precisely what I find in the third part of this dissertation. Viewed from the perspective of national legislative outcomes, as the responsiveness of Republicans to high income districts has increased, their responsiveness to the educated has decreased. That is, Republicans are increasingly successful in legislation for richer districts, and decreasingly successful in educated districts.

There are theoretical grounds to believe that education and income will have disparate effects on policymaking. While Republicans tend to win the votes of the rich, Democrats are known as the party of Ivy League intellectuals (Ansolabehere, Rodden, & Snyder, 2006; Gilens, 2012). Republicans are thought to be more responsive to economic interests of the rich, while Democrats are more responsive to the intellectual elite. Therefore Democrats should respond to the highly educated more than Republicans.

Yet it is also plausible that education and income have the same effect in other areas of politics. High education and high income are both indicators of achievement and engagement in society. Individuals engaged in society can be expected to be more engaged in politics and more likely to demand policy congruence from their representatives. There are many questions to be answered about when the effect of income and education should have similar impacts on politics, and when they should be different.

In practice, the lack of theoretical grounding for the independent impact of education relative to income on political outcomes means that the potential independent impact of education needs to be carefully explored whenever income is the primary explanatory variable. If both variables are included in the model and are statistically significant impact, then there are not interpretation issues due to the collinearity between these two variables. If both are included and neither are individually statistically significant, it is possible to draw the incorrect conclusion that they have no meaningful impact on the model. If education is omitted, it is possible to incorrectly assume the coefficient on income reflects the independent impact of income instead of the joint impact of income and education. In the worst case, this can lead to conclusions that income has a positive (negative) effect on the dependent variable, when in fact it has a negative (positive) and statistically significant impact. I show an example of this in voter turnout in a later section.

I illustrate the procedures outlined throughout this part of the dissertation using real world examples. These procedures proposed allow the joint and disparate effects of income and education to be untangled. I show that educational and economic characteristics of a district can have meaningful joint and independent effects on legislative outcomes, partisan preferences, and representational preferences.

10 EXAMPLE: LEGISLATIVE SUCCESS

The primary example I use theorizes a connection between district demographics, constituent preferences over policy, and legislative outcomes. The theoretical connection between district demographics and legislative activity lies in both representational and policy preferences. People prefer policy from their members of Congress when they are richer, as seen in Part 1. Republicans provide policy that is more congruent with the preferences of their constituents when those constituents have high incomes (Rhodes & Schaffner, 2017; Grossmann & Williams, 2018; Lax, Phillips, & Zelizer, 2018). Republicans are the party ideologically aligned with the interests of the rich, while Democrats are the party ideologically aligned with the interests of the educated and working class. Thus, Republican members of Congress who represent high income districts should find it easier to create successful legislation, as the legislation supported by their party and their ideology is congruent with the preferences of their constituents. Part 3 of this dissertation provides additional evidence for the mechanism behind this link.

The purpose of this section is to discuss and apply methods for detecting collinearity, implement the model interpretation guidelines laid out in this part of the dissertation, and present the visualization of the statistical significance and substantive effect of income and education in each model. I use two examples from the data to highlight model interpetation. In the first, drawn from data on Democratic legislative behavior from the 1980s, I show a situation where Case 1 applies. The graphics will clearly show why using the combined impact obscures the independent effect of both income and education. The second example is drawn from Republican legislative behavior in the 2000s. Here, Case 3 applies. Neither income nor education are independently significant, but they are jointly statistically significant. In this case, it is possible to draw meaningful inferences about the joint effect of income and education.

Additional examples provided in the next section will examine situations where Case 2 applies, as well as the different ways income and education can impact political outcomes.

Data

The primary independent variables are district income, district education, and a combined measure of district socioeconomic status. District income is measured as the percent of a legislator's district that earned over $75,000 per household per year in 2016 dollars. This captures the percent of the district that is high income. The cutoff for this is not always perfectly $75,000, as the value changes according to inflation and the income brackets used by the census. It is always in the range of $65,000-$75,000 in inflation adjusted 2016 dollars, and always falls above the median income of the nation at the time. The results are consistent across a variety of income measures. The secondary independent variable is district education, which is measured as the percent of the district with at least a bachelor's degree. All parts of the analyses separate Republicans from Democrats because the ideologies of each party create different kinds of responses across district economic and educational levels. In order to explicitly capture the joint impact of income and education, I find the principal component behind education and income and call it socioeconomic status (SES). The socioeconomic variable is a linear combination of education and income.

In addition to district income and education, I control for the percent of the district that is black.[10] I control for a variety of characteristics of the members of Congress: ideology as measured by DW-NOMINATE scores, their seniority, party affiliation, race, whether they serve on a powerful committee, and whether they serve as the chair of a committee. As seen in the appendix, education and income have the strongest correlation of any of the bivariate relationships in the model. As I show in the remainder of this analysis, this correlation substantially alters the interpretation of the results from the models I run.

The primary dependent variable is the number of bills a representative sponsors that are approved of in a House vote. This captures a measure of investment in

[10]As seen in the appendix, the distribution of the percent of the district that is black is strongly skewed, so I use the logarithm of the percent in my regression models.

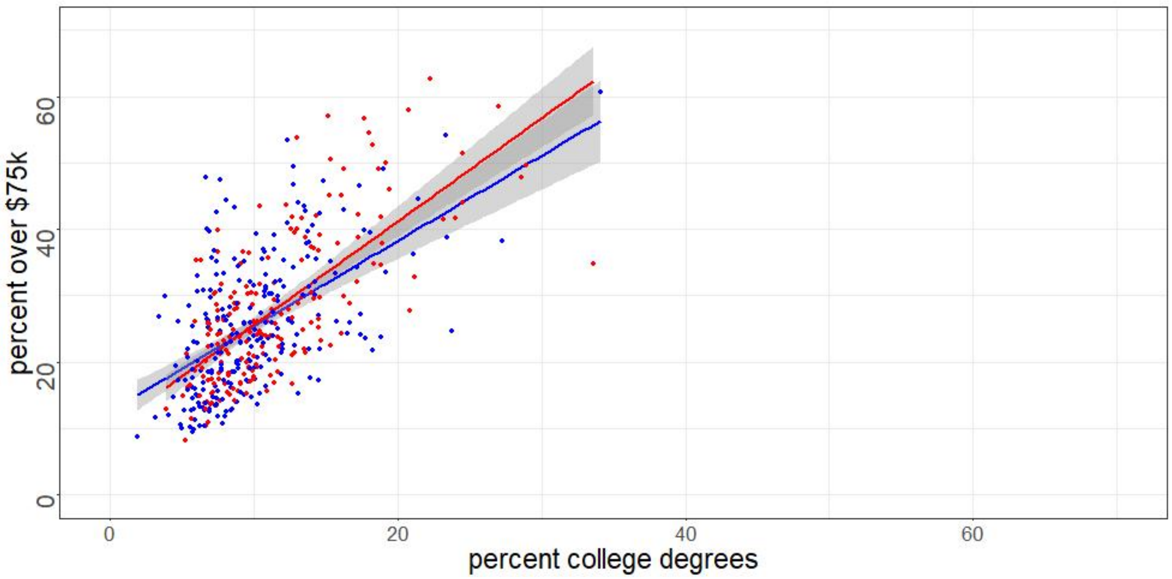
policy and legislative success. This is expected to vary in tandem with district income and education because these are correlated with how much constituents care about policy and congruence between the preferences of the representative and their constituents. Thus, district demographics serve as a proxy for the policy preferences of constituents.

Detecting Collinearity

As can be seen in Figure 10 and Figure 11, income and education are strongly correlated. The graph plots the district income and education for each representative in the 97th and 113th Congresses. Other Congresses have very similar correlations. Red dots represent districts represented by a Republican, and blue represents Democrats. The regression line is also depicted, in blue for Democrats and red for Republicans. Clearly, there is a correlation between district education and income for both Democrats and Republicans. This collinearity affects the ways omitted variable bias appears. As discussed in the prior section, it fortunately cannot artificially create statistically significant results in a model that includes all multicollinear variables.

These two Congresses were chosen because they show the underlying correlation between district education and income for two different cases for model interpretation. For Democrats between 1981 and 1986, Case 1 holds. For Republicans between 2011 and 2014, Case 3 holds. One pattern of note in these two graphs is that the range for spread of college degrees has substantially increased since 1981. The percent of a district with college degrees ranges between 15 and 65% today. In 1981 and 1982, is ranged between 3 and 35%. Another pattern of note is the fact that the tie between district income and educational attainment is higher today

Figure 10: Collinear Relationship between Education and Income, 1981-82



than it was in the past. These trends are consistent across all Congresses in the dataset.

The first example examines Democratic legislative behavior between 1981 and 1986. I will show how to determine that Case 1 applies to this example using diagnostics for multicollinearity. That is, the diagnostics will show that there is not much multicollinearity for this data. However, the key diagnostic is not whether the collinearity diagnostic tests indicate a problem, but whether the regression models fail to reject the null hypothesis because of multicollinearity.

Three diagnostic tools are commonly used to assess multicollinearity: the condition index, variance inflation factor, and a perturbation of the Y values to see if the coefficients substantially change. I will use a fourth, the bivariate plot of the simulated coefficients for income and education, which uses a similar method to the perturbation analysis but provides more detail. Table 3 and Table 4 depict the first three diagnostics for collinearity. They indicate that Case 1 applies. That is,

Table 3. Democratic 1981-1986 Collinearity Diagnostics: Condition Index and Variance Inflation Factor

Condition Index	Variance Decomposition Proportions							Weights
	intercept	income	education	year	conservativeness	seniority	percent black	
1.000	0.000	0.001	0.002	0.000	0.003	0.005	0.005	0.000
3.390	0.000	0.010	0.007	0.000	0.001	0.007	0.668	0.000
4.483	0.000	0.012	0.027	0.000	0.000	**0.850**	0.002	0.000
6.701	0.001	0.000	0.144	0.001	0.347	0.028	0.002	0.006
7.066	0.000	0.016	0.200	0.000	**0.645**	0.102	0.188	0.001
10.272	0.000	**0.830**	0.255	0.001	0.000	0.005	0.121	0.002
43.091	0.026	0.002	0.034	0.062	0.003	0.000	0.013	0.902
123.566	**0.973**	0.129	0.332	**0.937**	0.000	0.002	0.001	0.090

Democratic 1981-1986 Variance Inflation Factor					
income	education	year	conservativeness	seniority	percent black
1.320129	1.510961	1.288464	1.087829	1.021415	1.099920

Figure 11: Collinear Relationship between Education and Income, 2013-2014

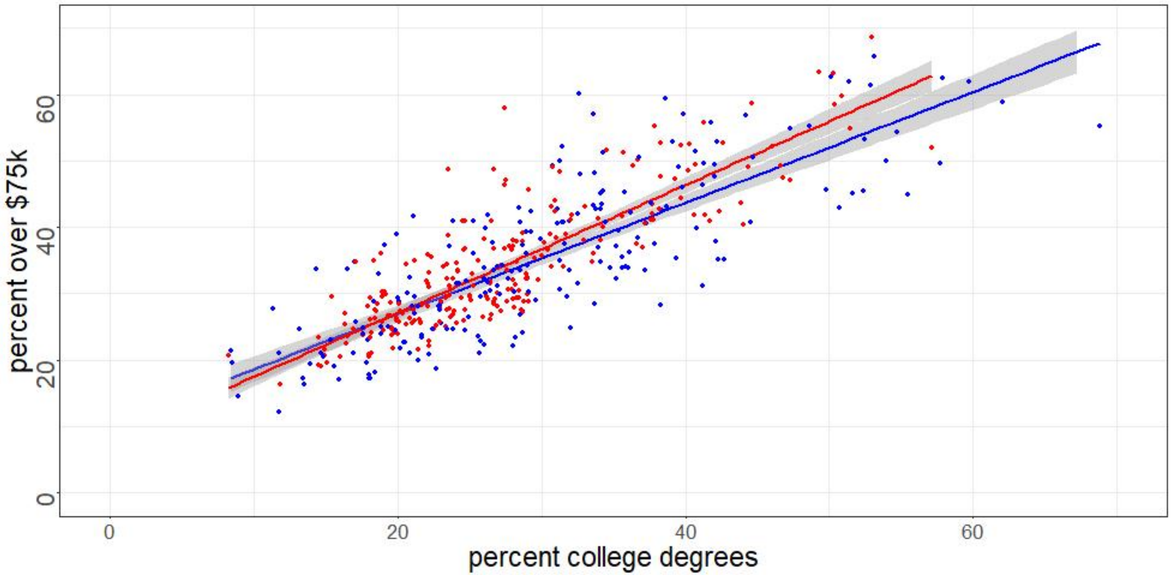


Table 4. Democratic 1981-1986 Collinearity Diagnostics: Perturbations

Impact of Perturbations on Coefficients				
	mean	std deviation	minimum coefficient	maximum coefficient
Intercept	1.843	0.074	1.674	2.035
income	-0.018	0.001	-0.016	-0.019
education	0.016	0.001	0.017	0.014
year	-0.023	0.001	-0.025	-0.022
conservativeness	-0.299	0.006	-0.312	-0.284
seniority	0.103	0.000	0.102	0.103
percent black	-0.003	0.000	-0.004	-0.003

the coefficients on income and education are both statistically significant in the full model that includes both.

The first diagnostic is the condition index, which is a measure of the degree to which the principal components of a variance-covariance matrix are unequal (Montgomery et al., 2012). Different sources provide slightly different cutoff values: Montgomery et al. (2012) sets the cutoff for high collinearity at 100, while the R package Perturb sets it at 30. Collinearity affects those variables that have a high variance decomposition proportion, set at 0.50 by Belsley (1991), or 50% of the variance inflation. These values are bolded in Table 3. While the condition index does rise above the cutoff values, it does not affect both education and income at the same time. Therefore, education and income do not have sufficient collinearity for Democrats between 1981 and 1986 to obscure the results. Instead, the high condition index is influenced by the intercept and the year. Collinearity between the intercept and year has no meaningful effect on this analysis.

The variance inflation factor is calculated from the correlation matrix (Montgomery et al., 2012). Again, the threshold for high collinearity varies, but it is common to set the cutoff at values higher than five or ten. None of the variance inflation factors exceed this cutoff. This supports the conclusion that the collinearity in the model does not meaningfully obscure the results.

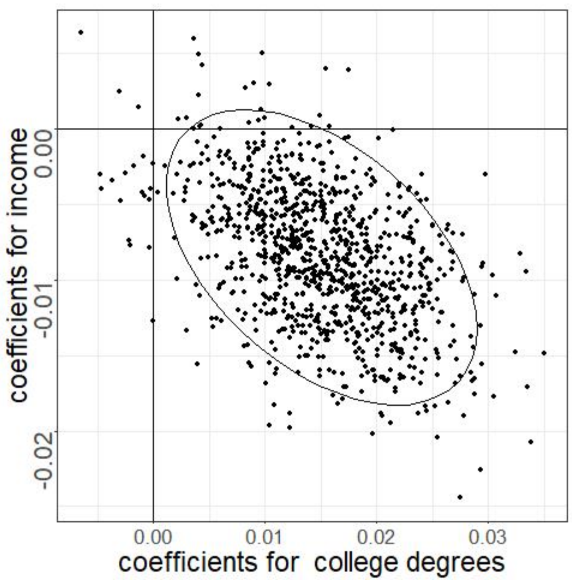
The most useful traditional diagnostic comes from the Perturb package in R. This package induces small changes to the variables to see if they unduly influence $\hat{\beta}$. It is considered the best test to see how much collinearity affects regression results. The results are presented in Table 4. It presents the lower bound of the 95% confidence interval for each coefficient, as well as the upper bound. If the coefficient does not change signs, it is a strong indication that collinearity does not affect the direction of the coefficients. For this group of observations, Democrats between 1981 and 1986, collinearity does not appear to influence the direction of the effects of income or education, nor meaningfully change the magnitude of the coefficient.

This perturbation can be expanded upon by plotting the simulated coefficients for education and income, as seen in Figure 12. These simulated coefficients are based on 1,000 runs of the the model in Equation 51

$$\text{legislative success} = \beta_1 * \text{income} + \beta_2 * \text{education} + \beta_3 * \text{controls} + \mu \tag{51}$$

where $\mu \sim N(0, \hat{\sigma})$ was independently drawn for each of the 1,000 runs. That is, each simulated set of coefficients is based on the same model with white noise added to the vector for legislative success. This is intended to give a general sense of the range of coefficients for β, but will not be precise.[11] That is, the average value for $\beta_{education}$

Figure 12: Simulated Coefficients for Education and Income, Democrats 1981-82



is positive, indicating that Democrats from highly educated districts passed more legislation than their equivalent counterparts representing less educated districts. The average value for β_{income} is negative, indicating that Democrats from high in-

[11]Beyond the fact that the estimates are necessarily imprecise by virtue of being estimates, the assumption that the noise is distributed normally is not exact. A negative binomial model assumes that the variance will be distributed as a function of both average number of bills that succeed, m, and the number of bills that do not succeed r, where $\sigma^2 = m + \frac{m^2}{r}$. Additionally, I truncated this normal distribution to prevent observations for Y that went below zero, as it is both impossible to have a negative amount of successful legislation and the negative binomial models do not allow for negative observations in Y.

come districts actually passed less successful legislation than their counterparts from districts with lower income levels, ceteris paribus. As I will show, because all else is not equal, this does not mean that the average highly income district tended to have a representative who was less legislatively successful.

The ellipse in Figure 12 is mildly, but not extremely, elongated. This indicates that for all that education and income are correlated, as seen in Figure 22 in the appendix, they are not substantially collinear in the context of all other variables.[12]

Overall, these diagnostics indicate that for Democrats between 1981 and 1986, Case 1 will apply. If there is an independent effect of income or education, it should appear in the full model. Yet these are merely indications, not conclusive.

Turning to an example of Case 3, I examine the same group of variables for Republicans between 2011 and 2014. As I will show, the diagnostics indicate that the collinearity between income and education bites in this case. Here, district income and education are both associated with increased legislative success. However, this time the effect of income (education) controlling for education (income) are artificially obscured by the inherent collinearity between the two variables.

The impact of collinearity can be seen in the diagnostics for Republicans between 2011 and 2014, shown in Table 5 and Table 6. While the variance inflation factor does not signal any major issues for the coefficients on income and education, the condition index and the perturbations do. For the condition index of 24.271, both education and income have very high variance decomposition values, well over the 0.50 or 50% threshold. This is reflected in the values of the coefficients when small perturbations are introduced, as described by Table 6. Here, the sign of the

[12]The top of the oval extends above the $\beta_{income} = 0$ intercept. Note that the shape of the oval will necessarily extend slightly beyonds 95% of the coefficients for income. Indeed, 4.1% of the simulated coefficients for income are positive, which is under twice the estimated p-value of 0.028 from the estimated coefficient for the observed values of legislative success in Model 1 of Table 7.

Table 5. Republican 2011-14 Collinearity Diagnostics: Condition Index and Variance Inflation Factor

Condition Index	Variance Decomposition Proportions							Weights
	intercept	income	education	year	conservativeness	seniority	percent black	
1.000	0.000	0.000	0.000	0.000	0.001	0.005	0.005	0.000
3.829	0.000	0.000	0.000	0.000	0.001	0.392	0.456	0.000
4.477	0.000	0.002	0.003	0.000	0.007	**0.533**	0.434	0.001
8.405	0.000	0.055	0.088	0.000	0.079	0.014	0.070	0.009
12.983	0.000	0.000	0.008	0.000	0.863	0.045	0.004	0.043
24.271	0.000	**0.936**	**0.898**	0.000	0.017	0.011	0.006	0.006
30.667	0.001	0.004	0.000	0.001	0.031	0.000	0.021	**0.933**
848.367	**0.999**	0.002	0.001	**0.999**	0.001	0.000	0.003	0.008

Variance Inflation Factor					
income	education	year	conservativeness	seniority	percent black
1.958373	1.939159	1.003454	1.023099	1.034675	1.010547

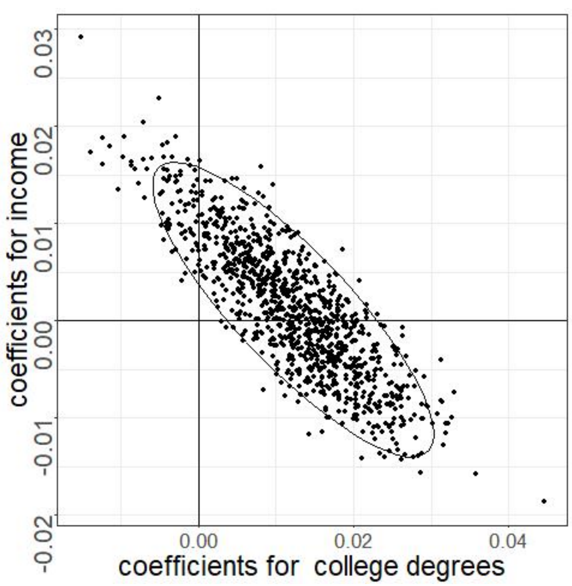
Table 6. Republican 2011-14 Collinearity Diagnostics: Perturbations

Impact of Perturbations on Coefficients				
	mean	std deviation	minimum coefficient	maximum coefficient
Intercept	-18.229	0.188	-18.632	-17.780
income	0.002	0.002	-0.005	0.008
education	0.013	0.003	0.006	0.021
year	0.166	0.002	0.162	0.169
conservativeness	-0.502	0.010	-0.526	-0.480
seniority	0.050	0.000	0.050	0.051
percent black	-0.025	0.000	-0.025	-0.024

coefficient on income in the model that has a control for education flips between -0.005 and 0.008.

This plot of the estimated coefficients from a similar perturbation are seen in Figure 13. As before, these are drawn from 1,000 simulations of the model in Equation 51, where μ is independently drawn for each simulation. As before, the

Figure 13: Simulated Coefficients for Education and Income, Republicans 2011-2014



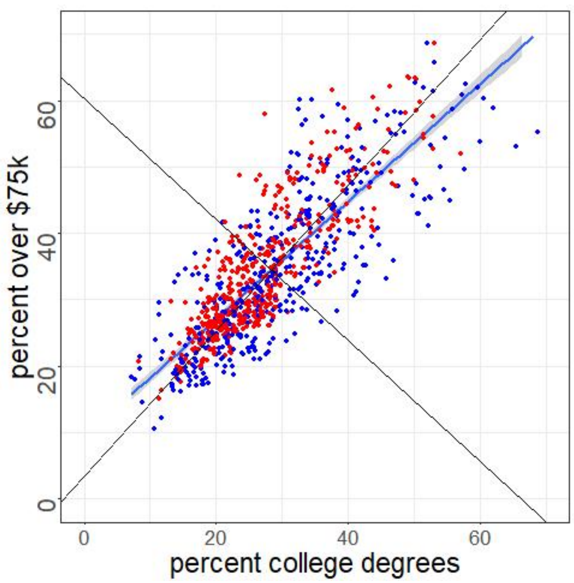
ellipse captures 95% of the estimated coefficients for income and education. This ellipse is much more elongated than that visible in Figure 12. That is, collinearity is leads to less certainty on the independent effect of income and education even though the joint effect is still positive and statistically significant. In this case, the estimated coefficients for both variables can be either negative or positive, as is the case for any coefficient that is not statistically significant. Thus, collinearity can be expected to effect the sign of the coefficient for income in addition to inflating the standard errors. The key fact is that their combined impact may still expected to lead to a statistically significant and substantively meaningful effect on the number of bills passed.

These results indicate that the coefficient on education, and possibly that for income, will not be statistically significant. Yet, as with the previous diagnostics, they are not conclusive. Whether multicollinearity bites depends, at least in part, on the luck of the draw for the data. The expected estimate of the coefficients are still centered at their true values, but will fall over the 95% confidence interval for the independent impact of at least one coefficient well over 95% of the time. The only conclusive test for whether multicollinearity will obscure the independent impact of variables is if the multicollinear variables are jointly significant but individually insignificant in the regression results.

10.1 Creating a Principal Component Called SES

Figure 11 demonstrated the correlation between Congressional districts with many highly educated constituents and those with many affluent constituents for the 112th and 113th Congress. Figure 14 shows the same plot overlaid with the new axes created by principal component analysis. As before, the regression line is plotted in blue. The first principal component is the black line with a positive slope. This reflects the length along with the greatest change in education and income jointly occurs. The secondary component, the black line with a negative slope, reflects the next highest amount of variation in the data. Note that the estimated regression line and the first principal component are similar, but not identical. The transformation

Figure 14: Correlation between District Income and Education 2013-2014



matrix for the matrix composed of X_{income} and $X_{education}$ for members of both parties in the same time period is found using the package "stats" in R, using the command prcomp():

$$V = \begin{bmatrix} a & -b \\ b & a \end{bmatrix} = \begin{bmatrix} 0.7389 & -0.6738 \\ 0.6738 & 0.7389 \end{bmatrix}$$

Principal component analysis involves a change in coordinate systems based on this transformation matrix V. The new coordinates for each observation are transformed to reflect the coordinate system defined by the two principal components. In this case, I define the first component to be socioeconomic status (SES), and the secondary components to be orthogonal SES. These are calculated based the values in the transformation matrix V, Equation 26, and Equation 27:

$$\text{SES} = -0.4456 + 0.7389 * \text{income} + 0.6738 * \text{education} \tag{52}$$
$$\text{orthogonal SES} = 2.304 - 0.6738 * \text{income} + 0.7389 * \text{education}.$$

These are used to calculate the new columns of observations, SES and orthogonal SES, that will be used in subsequent regression models.

Using the equations laid out in Equations 28 and 29, the new axes for the principal components coordinate system are defined as follows:

$$\text{income} = 60.30586 - 0.9119 * \text{education}$$
$$\text{income} = 3.419412 + 1.0966616 * \text{education}.$$

Figure 15: Correlation between District Income and Education 2013-2014

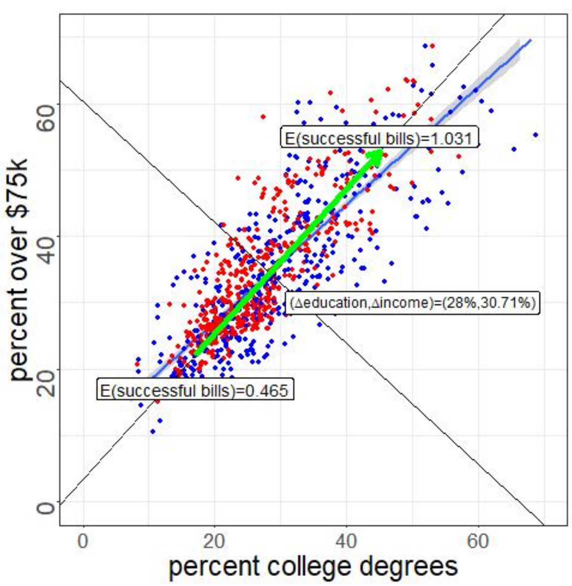


Figure 14 can be used to interpret the meaning of the effect of β_{SES}. It is the expected change in Y for a unit change along the new SES axis. Yet a unit change along the principal component is not intrinsically meaningful. To make the change calculated concrete, Figure 15 shows the change in successful legislation from a district near the bottom quartile of both the income and education distributions as compared to a district near the top for Republicans between 2013 and 2014, as calculated from the model presented in Table 8 in the next section. The difference of 0.566 is substantial considering the fact that the average Republican only sponsored 1.722 pieces of successful legislation in a given session of Congress.

Figure 16: Correlation between District Income and Education 2013-2014

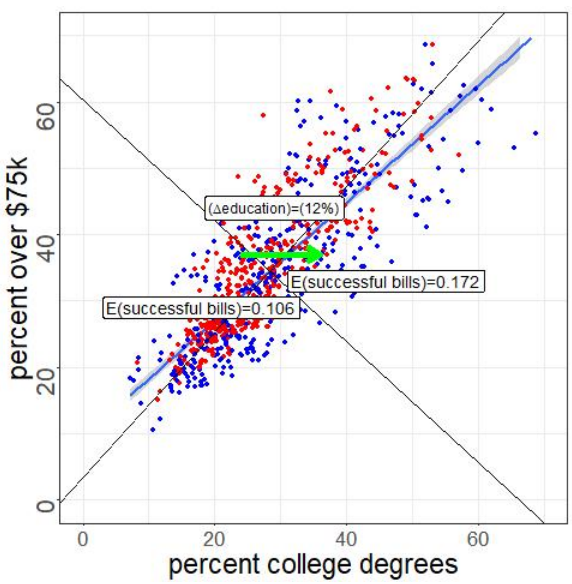


Figure 16 shows the limited nature of calculating the independent effect of education for Republicans. Note that this is based on coefficients that are not individually statistically significant, so this is a very rough guess. Yet it is still apparent that to compare a highly educated Republican district against a less educated Republican district, ceteris paribus, one can only compare districts with around a 12% difference in the percent of the district that has a college education. Note that in this direction, the effect of education appears to be minimal.

Of course, this is not always the case. I will show that for Democrats in Table 7 in the next section, the only meaningful effect is seen in a similar independent effect for education and income. Since the magnitudes of both effects will be similar but in opposite directions, their expected net effect washes out as one moves along the direction of the SES axis.

Model Interpretation

The goal is to understand the statistical significance and substantive impact of income and education. Because they are correlated, both their independent impacts and joint impacts are relevant. The four models that reveal these effects are:

1. Model 1 includes both income and education.

$$\text{legislative success} \sim \beta_1\text{*income}+\beta_2\text{* education} + \beta_3\text{*controls}$$

2. Model 2 includes income while omitting education.

$$\text{legislative success} \sim \beta_1^{\cancel{\text{edu.}}}\text{* income} + \beta_3^{\cancel{\text{edu.}}}\text{*controls}$$

3. Model 3 includes education while omitting income.

$$\text{legislative success} \sim \beta_2^{\cancel{\text{inc.}}}\text{* education} + \beta_3^{\cancel{\text{inc.}}}\text{*controls}$$

4. Model 4 replaces both income and education with a socioeconomic status variables created from income and education through principal component analysis.

$$\text{legislative success} \sim \beta_{SES}\text{* SES} + \beta_{orth}\text{* orthogonal} + \beta_3\text{*controls}$$

The diagnostics above indicated that the association between district demographics and legislation for Democrats between 1981 and 1986 would land in Case 1, where the full model will reveal statistically significant effects on income and education if they exist. They also indicated that a similar association for Republicans would land in Case 3, where the coefficients on income and education are jointly, but not individually, statistically significant. Model 1 is used to check this along with joint significance tests.

The results for Democrats presented in Table 7 show that the coefficients for income and education are statistically significant in Model 1. Yet the coefficient on SES in Model 4 is statistically insignificant, revealing that the interpretation of the coefficients on income and education in Model 1 should be treated with care.

Democrats who represented highly educated districts produced more policy, and they produces less when they represented high income districts. This is congruent with the modern policy platforms of the Democratic party. It is also congruent with modern voting behavior: the highly educated tend to vote Democratic, while the rich tend to vote Republican (Gelman, 2009; McCarty et al., 2006). Yet a district at the top end of the income and education spectrum will not have a Democratic member of Congress who appears to produce more legislation relative to a district at the bottom end of both income and education. This is because the effects of income and education are countervailing.

The results presented in Table 8 show that multicollinearity bites. As seen in Model 1, neither coefficient for education or income is statistically significant. Indeed, the standard error on income is substantially larger than the estimated coefficient. Yet they are jointly significant, as seen in both an F-test and the coefficients for the joint impacts visible in Models 2-4. This means that Case 3 applies,

Table 7. Collinearity Examples, Democrats 1981-86

	Model 1	Model 2	Model 3	Model 4
income	-0.015^{*}	-0.006		
	(0.006)	(0.005)		
education	0.025^{**}		0.014	
	(0.009)		(0.008)	
SES				0.006
				(0.005)
orthogonal				0.028^{**}
				(0.010)
conservativeness	-0.730^{*}	-0.820^{**}	-0.601^{*}	-0.730^{*}
	(0.305)	(0.307)	(0.299)	(0.305)
seniority	0.112^{***}	0.110^{***}	0.114^{***}	0.112^{***}
	(0.012)	(0.012)	(0.012)	(0.012)
percent black	-0.068^{*}	-0.061^{*}	-0.057	-0.068^{*}
	(0.031)	(0.031)	(0.031)	(0.031)
Black Caucus	-0.375	-0.420	-0.330	-0.375
	(0.216)	(0.217)	(0.216)	(0.216)
comm. chair	0.591^{***}	0.575^{***}	0.600^{***}	0.591^{***}
	(0.163)	(0.165)	(0.165)	(0.163)
powerful comm.	-0.673^{***}	-0.687^{***}	-0.678^{***}	-0.673^{***}
	(0.112)	(0.112)	(0.112)	(0.112)
Observations	768	768	768	768
Akaike Inf. Crit.	2,550.349	2,555.514	2,553.602	2,550.349

Note: Negative Binomials $^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$

Table 8. Collinearity Examples, Republicans 2011-14

	Model 1	Model 2	Model 3	Model 4
income	0.004	0.016^{***}		
	(0.009)	(0.005)		
education	0.016		0.020^{***}	
	(0.010)		(0.005)	
SES				0.014^{***}
				(0.004)
orthogonal				0.009
				(0.013)
conservativeness	-0.603^{*}	-0.534	-0.615^{*}	-0.603^{*}
	(0.280)	(0.278)	(0.279)	(0.280)
seniority	0.026^{*}	0.024^{*}	0.027^{*}	0.026^{*}
	(0.012)	(0.012)	(0.011)	(0.012)
percent black	-0.212^{***}	-0.207^{***}	-0.213^{***}	-0.212^{***}
	(0.052)	(0.053)	(0.052)	(0.052)
comm. chair	0.901^{***}	0.911^{***}	0.896^{***}	0.901^{***}
	(0.144)	(0.145)	(0.144)	(0.144)
powerful comm.	-0.231^{*}	-0.204	-0.237^{*}	-0.231^{*}
	(0.116)	(0.115)	(0.115)	(0.116)
Observations	480	480	480	480
Akaike Inf. Crit.	1,614.676	1,615.063	1,612.844	1,614.676

Note: Negative Binomials $^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$

and it is appropriate to use any of Models 2-4 to interpret the joint effect of income and education on legislative success.

The substantive interpretation of the results in Table 8 is that the highly educated and high income districts tend to have Republican representatives who are more legislatively successful between 2011 and 2014. It entails that Republicans have an easier time producing legislation when their constituents are some combination of highly educated and high income. The independent impacts of income and education are impossible to uncover with the given data. It is possible that with more data, both the independent impacts of income and education would gain statistical significance. Yet it is also possible that a latent variable captured by income and education is causing both high income districts, high education districts, and legislative productivity for Republicans. If that is the case, then they may not have a true independent impact, merely a joint impact.

The regression results presented provide sufficient information to interpret the models that uncover the joint and/or independent effects of income and education. Yet it requires examining information from multiple models and coefficients. A graphical solution to highlight the relevant information would help streamline model interpretation. This is what I present in the next subsection: a graphic that quickly reveals that statistical significance and size of each coefficient for income and education in the four models. This graphic not only helps with model interpretation in the face of multicollinearity, but it also highlights the substantive impact of both district education and income on legislative success.

Visualizing Effects of Collinearity

In order to allow the magnitudes to be easily compared between different variables, I focus on the expected change in Y values and develop a visualization to highlight these changes. I focus on the expected change in Y caused by a change in X_i instead of directly on the coefficient β. I compare the expected Y value for a high value of X_i against the expected Y value for a low value of X_i. I choose to use the third quartile of X_i as a proxy for a high value, and the first quartile as a proxy for a low value.[13] Using this method, the units being compared across different variables is the same: a change in the Y value. An added benefit is that this method highlights whether the β value is substantively meaningful.

This approach also ensures that I am always comparing apples to apples. While it is possible to plot the coefficients themselves on the same graph, the coefficients for one variable are likely to be drastically bigger or smaller that coefficients for the other variable because they are comparing different measurements. Comparing the estimated change in Y ensures that the magnitudes are always measured in the same units.

Finally, this approach highlights the fact that when the spread in the data of one group is different from the spread of the data in another, the quartile location will be different between the groups. Comparing top vs bottom quartiles makes it more explicit that the variance of each group will affect the calculated magnitude of the effect.

Using this calculation of the expected magnitude of the effect of each variable, I develop a visualization that allows the reader to quickly examine both the

[13]This approach is common. Yet as noted in prior sections, this may not always be the best method to compare the independent effect of each variable, as the spread of the observations for the conditional distribution is much different from the spread in the unconditional distribution. This dissertation leaves the solution of this problem to future work.

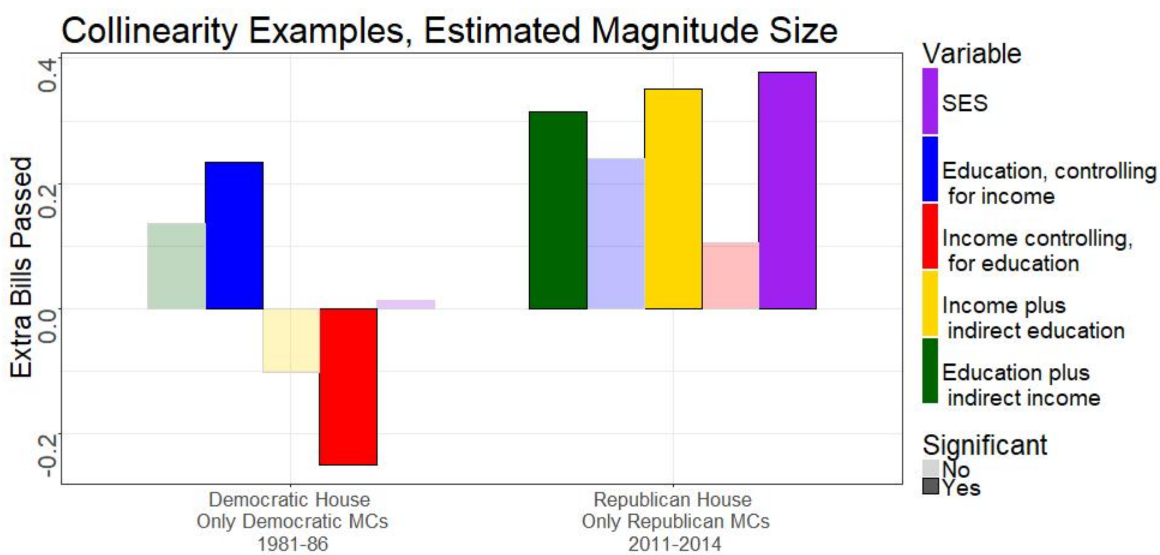
magnitude of the effect of the coefficient as well as the statistical significance. The statistical significance is relevant to determine how multicollinearity impacted the various models, and therefore which models to are appropriate. Showing the magnitude of the effect allows the reader to examine the impact of omitted variable bias, as well as the substantive effect of each explanatory variable.

Figure 17 graphically summarizes these two examples of the effects of collinearity. The y-axis shows the estimated number of extra bills passed by an individual member of Congress based on the socioeconomic characteristics of their district. This magnitude is estimated by comparing the number of successful bills sponsored by a legislator who represents the top quartile of the independent variable versus the number of successful bills sponsored by a legislator who is statistically identical except for representing a district at the bottom quartile. For example, the purple bar on the far right of Figure 17 shows that Republicans in the Republican controlled House between 2011 and 2014 who representative a district at the top quartile of the socioeconomic distribution sponsor 0.36 more successful bills than an equivalent representative from a district at the bottom quartile of the socioeconomic distribution. The effect is statistically significant, so the bar is bolded and outlined in black.

Each bar represents the magnitude for one of five coefficients derived from Models 1-4: the combined impact of district income and education channeled through education (green), education controlling for income (blue), the combined impact of district income and education channeled through income (yellow), income controlling for education (red), and a combined socioeconomic measure (purple). Note that the estimates based on the education variables are blue and green. The estimates based on the income variables are yellow and red. The socioeconomic variable is purple, which is the color wheel result from combining of blue and red.

Another benefit of this graphic is that the magnitudes of effects calculated from the coefficients for education and income in each model are comparable. In a typical regression table, coefficients often cannot be compared because they are measured in different units. For the graphic, I calculated the expected effect of each explanatory variable on the dependent variable. The result is that all magnitudes are measured in the same units as the dependent variable: number of successful bills sponsored. This allows the effect of district educational attainment to be directly compared against the effect of district income.[14]

Figure 17: Effects of Collinearity



The colors are faded when they are not statistically significant at the $\alpha = 0.05$ level, and bolded with a black outline when they are statistically significant. For example, for Democrats in the Democratic Houses between 1981 and 1986, the coefficient for education when income has been controlled for, in blue, is statistically

[14]As it so happens, both income and education are measured as a percentage of the district, so the units are the same. Yet it is unusual for variables to be measured in the same units.

significant. So is the coefficient for income when education is controlled for, in red. The other three bars are not statistically significant. Specifically, the coefficients for education without controlling for income (green), income without controlling for education (yellow), and overall socioeconomic status (purple) are not statistically significant. The opposite is true for Republicans in Republican Houses of 2011-2014.

Each group of five bars was chosen as a case study in the two main ways collinearity can effect the results. The left hand group of five bars shows the magnitude of the effect for each of the five coefficients for 1981-86, focusing on Democrats. The House was controlled by Democrats at this time. The right hand group of five bars shows the magnitudes of each coefficient for 2011-14, focusing on Republicans in the Republican controlled House.

One take away from Figure 17 is that for Democrats between 1981 and 1986, Case 1 applies. Multicollinearity does not bite and the independent impacts of district income (red) and education (blue) are statistically significant. As I discuss in the third part of this dissertation, the magnitudes described are also substantively meaningful because each Democratic member of Congress in this time period passed an average of only 1.72 bills per Congressional session. Thus, the blue and red bars represent the coefficients of Model 1. The faded purple bar shows that the countervailing effects of income and education lead to a coefficient on SES that is very close to zero and not statistically significant.

The faded bars on the left all represent the combined impact of income and education. They also provide an object lesson for why the coefficients in Models 2 and 3 must be interpreted as the joint effect of income and education. The faded yellow bar on the left is negative, but much closer to zero than the bold red bar. This indicates that when the joint impact of income and education is channeled through income, the combined effect appears to be negative but is statistically insignificant. Yet the joint effect of income and education, when channeled through education, is positive and not statistically significant. Clearly, the magnitude of the effect the combined measure represented by the green bar is lower than the independent effect of education because the coefficient for the combined measure incorporates the countervailing effect of the income.

Another take away from Figure 17 is that for Republicans between 2011 and 2014, Case 3 applies. Multicollinearity bites, and although income and education are jointly significant, it is impossible to distinguish their independent effects (faded red and blue). Yet the combined effect represented by the green, yellow, and purple bars is still substantively meaningful and statistically significant. The direction of the effect and statistical significance for all three are identical, and the magnitudes of effects are similar. Indeed, the magnitudes of the effects are larger than those evident for the independent impacts of income and education for Democrats on the left hand side.[15]

This graphic summarize a large amount of information to help understand non-orthogonal variables: statistical significance, direction of effects, and effect sizes for five coefficients derived from four models for each of two groups. Again, the goal is to capture the statistical significance and substantive importance of education and income. This part of the dissertation shows that this is possible so long as the joint

[15] A word of caution is relevant for comparing coefficients across groups. If the distribution of the number of bills passed is substantially different for Republicans between 2011 and 2014 versus Democrats in 1981-1986, then the differences between the groups will be not be due to the differences caused by income and education, but due to the differences due to the group. Fortunately, both time periods focus on the majority party, and members of the majority party have similar rates of legislative success across time and party. To use a separate example, it would not be meaningful to compare how weight changes the heart rate of a mouse as compared to how it changes the heart rate of a giraffe. Yet comparisons may still be possible for a different measure, such as the percentage change in heart rate.

impact of income and education is statistically significant. It demonstrates a method to uncover the substantive effect by appropriately interpreting coefficients based on whether they capture the independent or joint effects of income and education.

11 ADDITIONAL EXAMPLES of EDUCATION and INCOME in POLITICS

The remainder of this part of the dissertation will examine three other political outcomes in which the collinearity between education and income influences the interpretation of the coefficients in Models 1-4. These are the party elected to Congress based on district education and income levels, how conservative those members of Congress are, and voter turnout. There are some theoretical expectations through these examples, but the primary purpose is to show the application of the model interpretation and to demonstrate that failing to accurately account for the confounding influence of education can lead to incorrect inferences. Specifically, I show that

- Failing to include education in a model predicting whether Republicans are elected can lead to the incorrect inference that district income does not appear to change who is elected. The correct inference is that district income increases the probability of electing a Republican, but the combined impact of income and education has no statistically significant effect on which party is elected.
- Failing to include education in a model predicting the ideology of members of Congress can lead to the incorrect inference that district income is associated with more polarized members of Congress. The correct inference for is that highly educated districts are associated with more polarized members of

Congress in the time period examined, but the independent effect of income is not statistically significant.

- Failing to include education in a model predicting voter turnout can lead to the incorrect inference that counties with high incomes or low unemployment rates have lower levels of turnout. The correct inferences is that counties with high unemployment rates, and correspondingly low incomes, have higher levels of turnout.

Party Elected to Congress

There has been some debate over the influence of income on which party is elected to Congress. This section will show that if the effect of education is not controlled for, then it will appear as if income has no statistically significant impact on the party elected. That is, the countervailing independent effect of education is substantial enough shrink the combined effect of income and education down toward zero.

This example uses the same dataset examined in the prior section, but this time the focus is on which party is elected to Congress based on district demographics. In order to determine which model best captures the effect of district education and income, I present the results of Models 1-4 in Table 9 and Figure 18.

Model 1 shows that income (red bar) and education (blue bar) have statistically significant independent impacts on which party is elected to Congress. Thus, Case 1 applies.The magnitude of the coefficient is on the small side. A district at the top quartile of educational attainment (blue bar) is 1.6% more likely to elect a Democrat to Congress than an equivalent district at the bottom quartile of educational attainment. Meanwhile, a district at the top quartile of the district income (red bar) spectrum is 1.4% more likely to elect a Republican. This means that we

Table 9. Probability Republican Elected 2007-2010

	Model 1	Model 2	Model 3	Model 4
district income	0.051^{***}	0.012		
	(0.012)	(0.006)		
district education	-0.053^{***}		-0.006	
	(0.013)		(0.007)	
district SES				0.002
				(0.004)
percent black	-0.290^{***}	-0.315^{***}	-0.334^{***}	-0.327^{***}
	(0.071)	(0.070)	(0.070)	(0.070)
Observations	896	896	896	896
Akaike Inf. Crit.	1,156.620	1,171.687	1,174.016	1,174.582

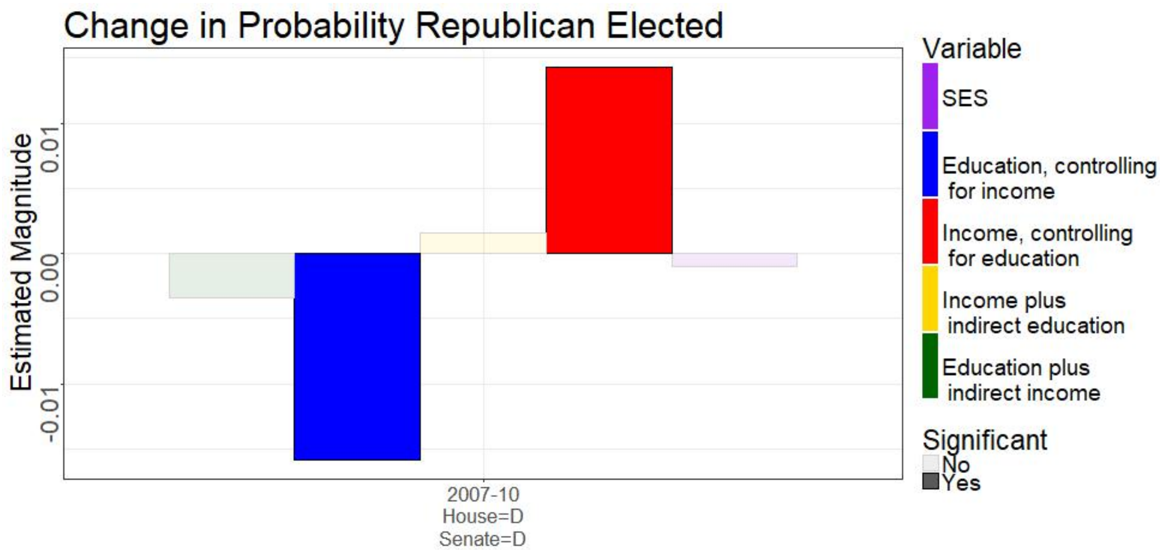
Note: Logistic regression *p<0.05; **p<0.01; ***p<0.001

could expect to see around just over one fewer Republicans elected from districts at the bottom end of the educational attainment spectrum, and a bit over one more elected from districts at the top end of the district income spectrum.

Yet although Model 1 best reflects the impacts of education and income, Models 2-4 still provide some interesting insights. They highlight the fact that the combined impact of income and education has no statistically significant effect on the party of elected to Congress. That is, the three measures of the joint impact of income and education (faded green, faded yellow, and faded purple) are not statistically significant. This shows that if education is excluded from the model, the true positive effect of income on the probability a Republican is elected will not be visible.

One additional result from Table 9 is that, unsurprisingly, districts with large numbers of black constituents are unlikely to elect a Republican. Finally, AIC indicates that the Model 1 is a more informative model.

Figure 18: Effect of District Demographics on Party Elected to Congress 2007-2010

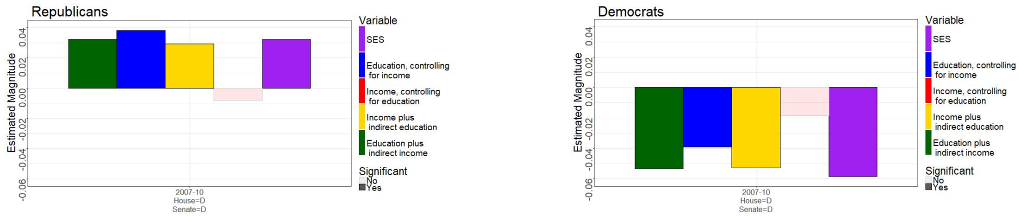


Ideology over Time

Another example of the effects of the collinearity between education and income comes from replicating a study on the influence of district income and education on a representative's conservativeness I examine how members of Congress vote on all bills by examining their first dimension DW-NOMINATE scores. DW-NOMINATE scores are a measure developed in Poole and Rosenthal (1997), and are commonly used to capture how conservative or liberal a member of Congress is.

This section replicates and expands on results from McCarty et al. (2006). As they show, district income and education have clearly distinguishable effects on the ideology of members of Congress. They also show that high income districts have become increasingly conservative over the past 60 years. By using the methods presented in this section of the dissertation, I show that this effect is due to Republicans taking control of Congress and because Republicans now lean more conservative when they represent highly educated districts.

Figure 19: Impact of Socioeconomic Characteristics on Conservativeness 2007-2010



(a) Democrats (b) Republicans

The dependent variable is the ideology of legislators, as measured by first dimension DW-NOMINATE scores. The more conservative a legislator, the closer his or her ideological score is to 1. The more liberal legislators have ideological scores close to -1. The economic and educational independent variables all range between 0 and 1, so their coefficients can be compared relatively directly.[16]

Before I examine the changing impact of district education and income over time, I will focus on a simpler case from 2007-2010. These results are presented in Figure 19. It is clear from the graphic that Case 2 applies: in Model 1, represented by the blue for the independent impact of education and faded red for the independent impact of income, only one is statistically significant. Case 2 indicates that either the full model, represented by the red and blue bars in the graphic, or the model that omits the variable for income, represented by the green bar, can be interpreted if care is used. If the model that omits income is selected, then the coefficient on the education variable (green) must be interpreted as the joint effect of income and education.

The graphics show that Republicans are more conservative (positive values) when they represent both highly educated and jointly highly educated and high

[16]The combined factor for socioeconomic status, SES, was normalized to range between 0 and 1. Education and income are both percentages that inherently are bounded between 0 and 1. Education reflects the percent of the district that has a college degree, and income reflects the percent of the district that earns over ~$75,000 in 2009 inflation adjusted dollars.

income districts. Democrats are more liberal under the same conditions. In other words, highly educated districts have more ideologically polarized representatives, as do districts that are both highly educated and high income. Nothing can be said about the independent impact of district income, as it is statistically insignificant in both models.

Note that it would be easy to choose the model that omits education (yellow), especially since education has not widely been theorized to be a more important factor in political outcomes than income. The graphic shows what happens if Model 2 is selected (yellow). The yellow bar shows that the combined impact of income and education is polarizing. Yet it would be easy to misinterpret this variable as the independent impact of income and conclude that high income districts have polarized representatives.

I now turn to examining the changing impact of district demographics across time. McCarty et al. (2006) show that connection between district income and legislator conservativeness has increased over the past 40 years. They argue that this is partially due to an increase in magnitude of the coefficient, but also because districts themselves are facing larger inequality.

Tables 10 and 11 show this is also due to the fact that Republicans are now more conservative when they represent districts with high socioeconomic status, not because all legislators have become more conservative when representing districts with high socioeconomic status. Democrats are still more liberal when they represent high socioeconomic status districts, just as they were in the 1970s and 1980s.[17]

Figure 20 highlights the changing effects of income, socioeconomic status, and education. As in the prior graphics, each bar in the graphics represent one measure

[17] The effect has declined somewhat for Democrats, but it is still strongly negative.

Table 10. Republican Ideology by District Demographics 1972-2014

	Model 1	Model 2	Model 3	Model 4
income	0.014*	−0.015***		
	(0.006)	(0.004)		
income*year	−0.0001*	0.0001***		
	(0.0001)	(0.00004)		
education	−0.050***		−0.036***	
	(0.008)		(0.005)	
education*year	0.0005***		0.0003***	
	(0.0001)		(0.0001)	
SES				−0.015***
				(0.003)
SES*year				0.0001***
				(0.00003)
year	0.024***	0.025***	0.023***	0.024***
	(0.001)	(0.001)	(0.001)	(0.001)
majority	0.015*	0.022**	0.020**	0.022**
	(0.007)	(0.007)	(0.007)	(0.007)
seniority	−0.008***	−0.008***	−0.008***	−0.008***
	(0.001)	(0.001)	(0.001)	(0.001)
percent black	−0.010***	−0.010***	−0.010***	−0.010***
	(0.002)	(0.002)	(0.002)	(0.002)
comm. chair	−0.001	−0.003	−0.002	−0.003
	(0.014)	(0.014)	(0.014)	(0.014)
powerful comm.	−0.023***	−0.021***	−0.022***	−0.022***
	(0.006)	(0.006)	(0.006)	(0.006)
Observations	4,134	4,134	4,134	4,134
Akaike Inf. Crit.	-3,205.318	-3,168.685	-3,202.619	-3,185.130

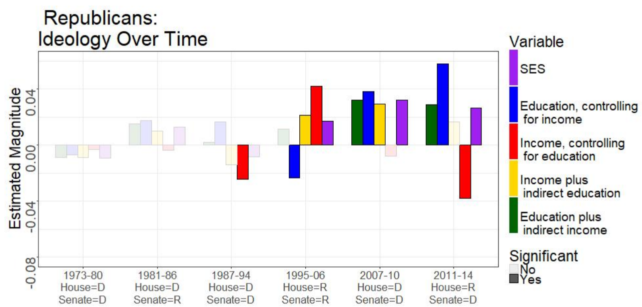
Note: OLS, *p<0.05; **p<0.01; ***p<0.001

Table 11. Democratic Ideology by District Demographics 1972-2014

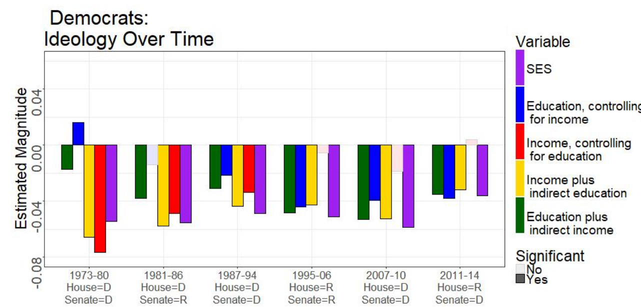
	Model 1	Model 2	Model 3	Model 4
income	-0.037^{***}	-0.022^{***}		
	(0.004)	(0.003)		
income*year	0.0003^{***}	0.0002^{***}		
	(0.00004)	(0.00003)		
education	0.015^{**}		-0.010^{*}	
	(0.005)		(0.004)	
education*year	-0.0002^{**}		0.0001	
	(0.0001)		(0.00004)	
SES				-0.016^{***}
				(0.002)
SES*year				0.0001^{***}
				(0.00002)
year	-0.007^{***}	-0.007^{***}	-0.001	-0.004^{***}
	(0.001)	(0.001)	(0.001)	(0.001)
majority	-0.031^{***}	-0.043^{***}	-0.022^{***}	-0.033^{***}
	(0.006)	(0.006)	(0.006)	(0.005)
seniority	-0.005^{***}	-0.005^{***}	-0.005^{***}	-0.005^{***}
	(0.0005)	(0.0005)	(0.001)	(0.0005)
percent black	0.017^{***}	0.017^{***}	0.022^{***}	0.019^{***}
	(0.002)	(0.002)	(0.002)	(0.002)
Black Caucus	-1.653^{***}	-1.611^{***}	-1.447^{***}	-1.530^{***}
	(0.113)	(0.112)	(0.113)	(0.111)
comm. chair	-0.024^{**}	-0.024^{**}	-0.023^{*}	-0.024^{**}
	(0.009)	(0.009)	(0.009)	(0.009)
powerful comm.	-0.032^{***}	-0.034^{***}	-0.032^{***}	-0.032^{***}
	(0.005)	(0.005)	(0.005)	(0.005)
Black Caucus*year	0.014^{***}	0.013^{***}	0.012^{***}	0.012^{***}
	(0.001)	(0.001)	(0.001)	(0.001)
Observations	5,060	5,060	5,060	5,060
Akaike Inf. Crit.	-5,717.505	-5,661.439	-5,527.723	-5,684.847

Note: OLS, *p<0.05; **p<0.01; ***p<0.001

Figure 20: Impact of Socioeconomic Characteristics on Conservativeness



(a) Republicans



(b) Democrats

of socioeconomic status from a regression that looked at that group of legislators. This time, the height of the bar reflects the size of the coefficient instead of the magnitude of the effect. For example, the green bar for Democrats between 1973 and 1980 shows the coefficient for education on Democrats ideology in that time period, controlling for the all non-economic or educational variable in Table 11.[18] So between 1973 and 1980, Democrats who represented highly educated districts were more liberal.

The impact of district demographics on Democratic ideology is highly consistent across time. Democrats who represent districts with many constituents who are of high socioeconomic status are consistently more liberal, whether socioeconomic status is measured by income, education, or both. The one exception is the effect of education when controlling for income in 1973-1980. For that time period and that measure, Democrats were more conservative when they represented districts with high education levels relative to their income level. The effects are almost always statistically significant. The increasing impact of income on conservative ideology is not due to Democratic behavior.

Once again reflecting the trends in legislative success, the impact of district demographics on Republican ideology is less consistent over time. In fact, prior to the 1994 Republican take over of the House, district socioeconomics had almost no impact on Republican ideology. The fact that high income districts overall create more conservative legislators is entirely driven by the changes over time for Republicans, as well as their newfound control over the House of Representatives. Note, however, that while income (yellow bar), education (green bar), and the combined socioeconomic status variable (purple bar) are always positive and usually statisti-

[18]Majority party is also not controlled for, as this time period was always controlled by Democrats.

cally significant, the effect sometimes becomes negative when education or income are controlled for. Between 1995 and 2006, the independent effect of education (blue bar) for Republican ideology was negative. That is, for two districts with similar numbers of high income inhabitants, the district with more educated inhabitants would be expected to have a legislator who was more liberal. Similarly, between 1987 and 1994, and between 2011 and 2014, the independent effect of income (red bar) was associated with more liberal Republican legislators. This discrepancy is currently unexplained, and merits future investigation.

Overall, the increasing association between district socioeconomics and legislator ideology is driven by Republicans, not Democrats. We can see this in Figure 20. This graphic highlights the differences between Democrats and Republicans. This is a very similar pattern to the one revealed in the analysis of legislative success, and is consistent with the theory presented by (Barker & Carman, 2012) and the voting patterns described by (Gelman, 2009). Republicans have changed their ideological grounding, particularly since 1990, and that appears to be reflected in the demographic ties to how they vote and how they create legislation.

Note that yet again the disparate effects of income and education for Republican ideology are highly relevant. Generally, Republicans are more conservative when they represent socioeconomically elite districts. Yet the opposite effect can be obscured if they are not examined both together and separately. For example, look at Democrats in a Republican House, as seen in Figure **??**. It is impossible to tell whether the effect of income when controlling for education is obscured by collinearity or if the apparent effect of income is entirely due to the collinearity with education. Yet we can see that overall socioeconomic status, as well the stand-alone impact of income, are both tied to more liberal Democrats. As another example, focus on the impact of income for Republicans in a Democratic House. Here, the independent effect of income once education is controlled for shows a correlation with more liberal members of Congress. Yet if income is looked at alone, without controlling for education, it could be misinterpreted to indicate that the district income is correlated with more conservative Republicans.

A few other results from these tables stand out. The first is that Republicans with large numbers of black constituents tend to be more liberal, while Democrats with large numbers of black constituents tend to be more conservative. This is an unexpected finding. Future work could examine whether this effect is due to Southern Democrats prior to the Republican revolution of 1994. However, a black Democratic member of Congress is, as expected, more liberal than a white Democrat.[19]

In general, Republicans have become more conservative over time. The trend for Democrats is less clear, as it depends highly on which demographic variables are controlled for. Intriguingly, when facing a Republican majority Democrats become more liberal while Republicans become more conservative.

Voter Turnout

The confounding effect of education is also apparent in a replication of the results presented in Burden and Wichowsky (2014). They claim that a common understanding used to be that counties with high levels of unemployment would have lower turnout rates. They make the argument that, once adequate controls have been included in the model, counties with higher turnout rates in fact have lower levels of unemployment. I will demonstrate that a key confounding variable is the educational attainment of the county.

[19] This result may be strongly influenced by whether the number of black constituents in the district is controlled for.

This is another example where Case 1 applies. That is, when the full model is used education and unemployment both have a statistically significant and countervailing effects. I will show that the same holds for an analysis where income is a primary explanatory variable instead of unemployment, and unemployment is omitted from the model.

The following two tables demonstrate the effect of collinearity between income and education in voter turnout rates. The analysis in the original paper refutes conventional wisdom that county unemployment rates have a negative relationship to voter turnout, and provides a theoretical grounding for the opposite effect. I replicated these results and focused on the impact of education, unemployment, and and income.

Table 12 reveals one potential origin of conventional wisdom. Namely, when education is not controlled for as in Model 4, it appears that there is a statistically significant negative correlation between unemployment rates and voter turnout. However, controlling for education reveals that this effect was driven by education rates, and once education is controlled for, as in Models 1-3, the effect of unemployment reverses while remaining statistically significant. That is, counties with high unemployment rates do have lower turnout for voting, but this effect is driven by the fact that counties with low high school graduation rates have both high unemployment rates and low turnout rates. Low educational attainment drives low voter turnout, while high rates of unemployment help increase turnout once educational attainment has been controlled for.

Of note, the effect of the collinearity between education and unemployment is not explicitly addressed in Burden and Wichowsky (2014). They acknowledge that their findings oppose tradition wisdom, but do not explain that traditional wisdom was confounded by omitting the influence of education. No other variable reversed the effect of county unemployment when omitted from the model; for this analysis, the most important collinear variable is education.

Table 12. Voter Turnout by Unemployment, Education, Gubenatorial Election (1980-2008)

	Voter Turnout			
	Full	(2)	(3)	(4)
County unemployment	0.146***	0.187***	0.191***	−0.290***
	(0.019)	(0.018)	(0.019)	(0.019)
High school graduation	1.506***	4.475***	4.510***	
	(0.107)	(0.060)	(0.060)	
Concurrent gubernatorial race	5.328***	1.747***		
	(0.299)	(0.141)		
State unemployment	0.510***			
	(0.035)			
Percent black	0.031			
	(0.017)			
Median income	−0.389			
	(0.215)			
Competitive presidential race	0.010***			
	(0.002)			
Concurrent senatorial race	0.634***			
	(0.060)			
Year fixed effects	Yes	No	No	No
County fixed effects	Yes	No	No	No
AIC	167704.3	203835.5	203986.1	209162.5
Observations	27,899	27,901	27,901	27,901
Note:				*p<0.05; **p<0.01; ***p<0.001

Additional analyses show that this effect is primarily created by counties with the lowest rates of educational attainment, as shown in Table 13. The interaction effect between unemployment and educational attainment is statistically significant.[20] Counties with the lowest education levels produce higher turnout when they face

[20]Model with the interaction effect is not shown.

high unemployment rates. The effect of unemployment disappears for counties in the top half of educational attainment.[21]

Yet their primary independent variable is unemployment, not income. To demonstrate that the effects are similar to the results in this dissertation, these results must be very similar when the analysis focuses on income instead of unemployment. In this case, because median income and county unemployment should produce similar effects on turnout rates, median income and unemployment rates should produce interchangeable statistical results. Table **??** shows that this effect holds when unemployment rates are omitted, instead focusing primarily on income. Namely, low income counties have overall lower turnout rates, but only when education is not accounted for. Low income counties of similar education levels have higher than expected turnout rates.

As the results from Table 12 show, when both income and unemployment are included in the model, only unemployment shows up as statistically significant. That is, unemployment does a better job of explaining turnout than does median income. In more precise terms, unemployment captures almost all of the variance in voter turnout that would otherwise be attributed to median income. Thus, for Burden and Wichowsky, unemployment is a more important explanatory variable than is median income.

[21]Note that multiple variables change magnitude substantially based on educational attainment. County unemployment and median income both become smaller as educational attainment goes up. Larger black populations are associated with reduced turnout in low education counties, and with increased turnout in high education counties. Competitive presidential races are associated with lower voter turnout in low education counties, but are associated with higher voter turnout in high education counties. One possible theory that explains this rests on the educational attainment of Democrats versus Republicans. Highly educated counties will tend to be more Democratic. Democrats tend to promote the participation of those who are poorer and minorities, so highly educated Democratic districts should see an increase in voter turnout when they have more poor and minority members. Yet Republicans tend to fan the anger of unemployed white males, so counties with high unemployment, particularly white male unemployment, should see higher levels of turnout relative to other similar counties.

Table 13. Voter Turnout by Educational Attainment

	Turnout based on county educational attainment quartiles			
	Lowest quartile	2nd lowest quartile	2nd highest quartile	Highest quartile
County unemployment	0.173***	0.109**	0.052	0.085
	(0.032)	(0.042)	(0.041)	(0.045)
High school graduation	2.169***	1.322***	1.811***	0.983**
	(0.270)	(0.384)	(0.438)	(0.302)
Concurrent gubernatorial race	4.755***	5.197***	4.387***	3.516***
	(0.569)	(0.553)	(0.593)	(0.794)
State unemployment	0.646***	0.708***	0.379***	0.331***
	(0.075)	(0.076)	(0.064)	(0.067)
Percent black	−0.037	−0.181***	0.281***	0.420***
	(0.033)	(0.042)	(0.037)	(0.040)
Median income	−2.283***	−2.610***	−1.501**	0.421
	(0.688)	(0.611)	(0.457)	(0.348)
Competitive presidential race	−0.010*	−0.0002	0.018***	0.027***
	(0.005)	(0.005)	(0.004)	(0.004)
Concurrent senatorial race	0.651***	0.672***	0.540***	0.414***
	(0.138)	(0.125)	(0.098)	(0.105)
Year fixed effects	Yes	Yes	Yes	Yes
County fixed effects	Yes	Yes	Yes	Yes
AIC	43092.7	42571.7	39231	40248.8
Observations	6,979	6,975	6,978	6,979

Note: *p<0.05; **p<0.01; ***p<0.001

Table 14. Voter Turnout by Income, Education, Gubenatorial Election (1980-2008)

	Model 1	Model 2	Model 3	Model 4
Median income	-5.305^{***}	1.183^{***}		
	(0.141)	(0.117)		
High school graduation	5.328^{***}		3.514^{***}	
	(0.074)		(0.058)	
SES				0.175^{***}
				(0.004)
Orthogonal SES				3.388^{***}
				(0.067)
Gubernatorial cycle	0.843^{***}	1.292^{***}	1.220^{***}	1.039^{***}
	(0.133)	(0.145)	(0.136)	(0.136)
Percent black	-0.133^{***}	-0.239^{***}	-0.148^{***}	-0.150^{***}
	(0.004)	(0.004)	(0.004)	(0.004)
Presidential cycle	-0.018^{***}	-0.017^{***}	-0.027^{***}	-0.023^{***}
	(0.003)	(0.004)	(0.003)	(0.003)
Senatorial cycle	0.430^{***}	0.446^{***}	0.407^{***}	0.292^{*}
	(0.112)	(0.122)	(0.115)	(0.114)
Year fixed effects	Yes	Yes	Yes	Yes
County fixed effects	No	No	No	No
Observations	27,921	27,921	27,921	27,921
Akaike Inf. Crit.	200,327.300	205,028.900	201,716.200	201,384.000

Note: OLS *p<0.05; **p<0.01; ***p<0.001

The results in Burden and Wichowsky point to yet another example of income related variables creating opposing political effects to education. I suggest that income and education frequently create opposing effects in politics. Any time that the impacts of education and income oppose each other, omitting one can dramatically obscure the independent effect of the other (Clarke, Kenkel, & Rueda, 2016).

12 CONCLUSION

Highly correlated explanatory variables can confound the interpretation of statistical models if not appropriately accounted for. One the one hand, omitting a highly correlated variable can substantially change the interpretation of the retained correlated variables. On the other hand, including all correlated and collinear variables will inflate the standard errors on the independent effects of each collinear variable. Multiple techniques have been proposed to account for these problems. I propose a solution that involves carefully considering all combinations of the collinear variables in models.

I argue that this model interpretation guideline will help scholars avoid interpretation problems when faced with multicollinearity. A model that has jointly significant collinear variables with no independent significance can still be interpreted to reveal the effect of their joint impact. The solution is to use principal components to combine the impact of the joint variables. A model that omits a highly correlated variable must be interpreted to be the combined effect of the omitted and included variables whenever it is known that the omitted variable has a high correlation with the included variable. This does not account for all the unknown correlated variables excluded from the model, but I argue that when we know about the potential for a confounded model, we should be extra careful about model interpretation.

In the process of demonstrating the justification for this interpretation of correlated variables, I highlight the importance of accounting for education in models of political outcomes. Whether the model used examines the joint impact of education and income, or the independent impact of education and income, inferential errors are possible. For example, the analysis of the impact of district income and education on Republican legislative success shows that Case 3 applies. This case is prone to the inferential error that income and education have no statistically significant effect on the model. Yet the correct inference is that education and income have a joint effect that is statistically significant. The magnitude of that joint effect is easily calculated.

On the other hand, the analysis of the impact of county income on voter turnout shows the danger of analyzing the joint impact of income and education if desired outcome is to understand the independent impact of income. Specifically, without accounting for education, the effect of county income appears to be positive. This fits easily with theories of income and voter turnout, but is not accurate. The correct interpretation of that effect is that the joint effect of income and education is positive. Yet closer examination shows that the independent effect of income is negative. That is, counties with more high income individuals have statistically significantly lower voter turnout rates. Burden and Wichowsky (2014) show that this is because counties with lower incomes have higher unemployment rates, and high unemployment rates are associated with higher voter turnout. I show that this effect is particularly strong for low education districts, which are more likely to be blue collar and rural.

The examples used throughout only involve two collinear variables. The next step would be to develop a method for interpreting more than two collinear variables. The number of combinations of multicollinear variables that would have to be examined would get large quickly, as the foundation for this model interpretation was an examination of one model for each possible combination of the two collinear variables.

References

Ansolabehere, S., Rodden, J., & Snyder, J. M. (2006). Purple America. *The Journal of Economic Perspectives*, *20*(2), 97–118.

Arceneaux, K., & Huber, G. A. (2007). What to do (and not do) with multicollinearity in state politics research. *State Politics & Policy Quarterly*, *7*(1), 81–101.

Barker, D. C., & Carman, C. J. (2012). *Representing red and blue: How the culture wars change the way citizens speak and politicians listen.* Oxford University Press.

Belsley, D. A. (1991). *Conditioning diagnostics: Collinearity and weak data in regression* (No. 519.536 B452). Wiley.

Bonica, A. (2014). Mapping the ideological marketplace. *American Journal of Political Science*, *58*(2), 367–386.

Burden, B. C., & Wichowsky, A. (2014). Economic discontent as a mobilizer: Unemployment and voter turnout. *The Journal of Politics*, *76*(4), 887–898.

Clarke, K. A. (2005). The phantom menace: Omitted variable bias in econometric research. *Conflict management and peace science*, *22*(4), 341–352.

Clarke, K. A., Kenkel, B., & Rueda, M. R. (2016). Omitted variables, countervailing effects, and the possibility of overadjustment. *Political Science Research and Methods*, 1–12.

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., … others (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, *36*(1), 27–46.

Farrar, D. E., & Glauber, R. R. (1967). Multicollinearity in regression analysis: the problem revisited. *The Review of Economic and Statistics*, 92–107.

Gelman, A. (2009). *Red state, blue state, rich state, poor state: Why Americans vote the way they do*. Princeton University Press.

Gilens, M. (2012). Under the influence. *Boston Review, July–August*.

Graham, M. H. (2003). Confronting multicollinearity in ecological multiple regression. *Ecology*, *84*(11), 2809–2815.

Greene, W. H. (2003). *Econometric analysis*. Pearson Education India.

Grossmann, M., & Williams, I. (2018). *Oligarchy or class war? Political parties and interest groups in unequal public influence on policy adoption.* (Working paper)

Jolliffe, I. (2011). Principal component analysis. In *International encyclopedia of statistical science* (pp. 1094–1096). Springer.

King, G. (1986). How not to lie with statistics: Avoiding common mistakes in quantitative political science. *American Journal of Political Science*, 666–687.

Lax, J., Phillips, J., & Zelizer, A. (2018). *The party or the purse? Unequal representation in the US Senate.* (Working paper)

Martin, A. D., & Quinn, K. M. (2002). Dynamic ideal point estimation via markov chain monte carlo for the us supreme court, 1953–1999. *Political Analysis*, *10*(2), 134–153.

Massy, W. F. (1965). Principal components regression in exploratory statistical research. *Journal of the American Statistical Association*, *60*(309), 234–256.

McCarty, N., Poole, K. T., & Rosenthal, H. (2006). *Polarized america: The dance of ideology and unequal riches (vol. 5).* MIT Press.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to linear regression analysis* (Vol. 821). John Wiley & Sons.

O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity*, *41*(5), 673–690.

Oster, E. (2017). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, 1–18.

Park, T., & Casella, G. (2008). The bayesian lasso. *Journal of the American Statistical Association*, *103*(482), 681–686.

Poole, K. T., & Rosenthal, H. (1997). *Congress: A political-economic history of roll call voting*. Oxford University Press on Demand.

Rhodes, J. H., & Schaffner, B. F. (2017). Testing models of unequal representation: Democratic populists and Republican oligarchs? *Quarterly Journal of Political Science*, *12*(2), 185–204.

Ridge regression. (n.d.). National Council for the Social Studies. Retrieved from `http://ncss.wpengine.netdna-cdn.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Ridge_Regression.pdf` (Accessed: 2018-06-14)

Russell, R., & Startup, M. (1986). Month of birth and academic achievement. *Personality and Individual Differences*, *7*(6), 839–846.

York, R. (2012). Residualization is not the answer: Rethinking how to address multicollinearity. *Social science research*, *41*(6), 1379–1386.

13 APPENDIX

13.1 Matrix of Scatter plots for District Demographics and Legislator Characteristics

Figure 21: Correlations between Independent Variables for Districts Prior to 1982

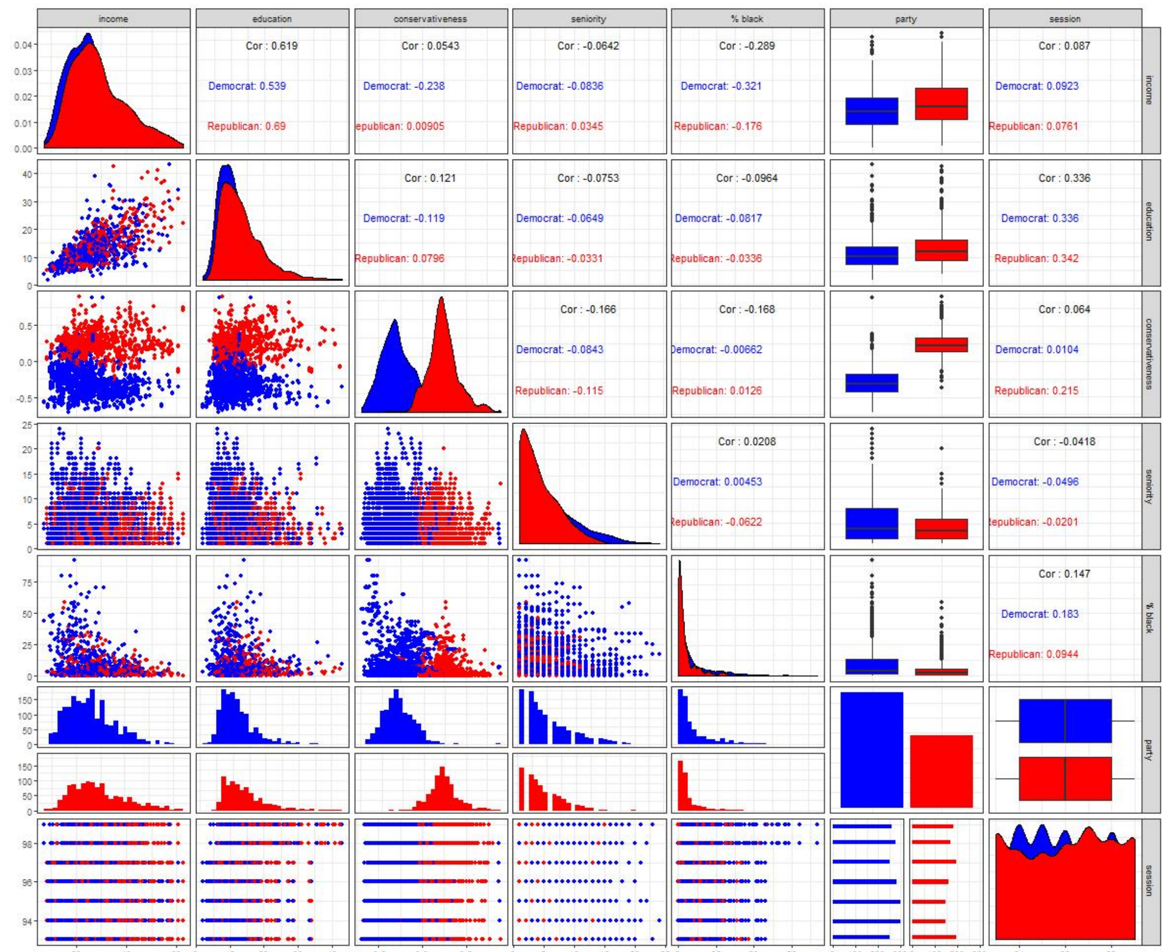
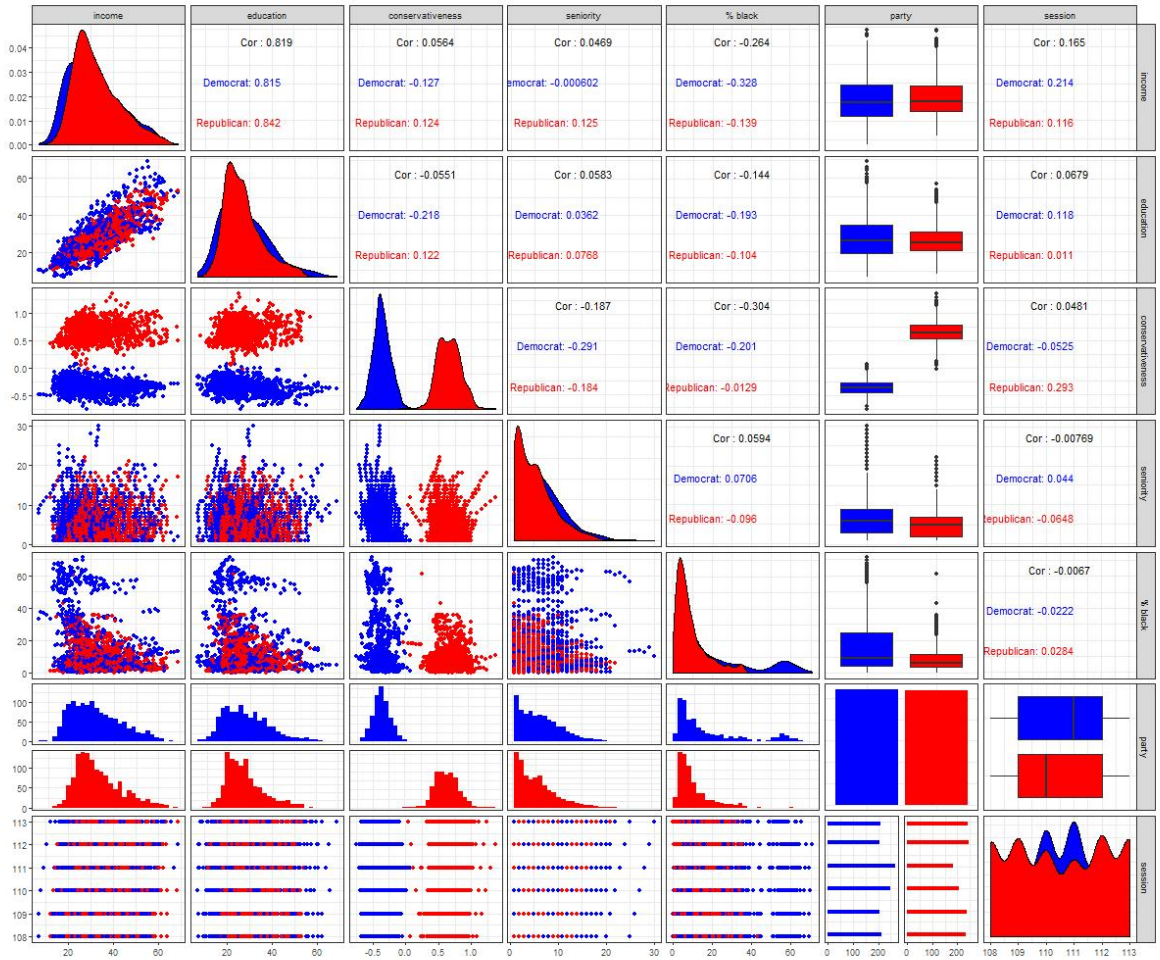


Figure 22: Correlations between Independent Variables for Districts After 2000



13.2 Deriving the Constant Term in the Principal Component Transformation

The calculation of the constant for the principal component axes relies on the transformation of the constant term in X. In order to rotate the matrix X_{23}, it must be centered. This means that the mean of each column is subtracted out. This is performed by multiplying

$$(X_0, X_2, X_3)\begin{bmatrix}1 & -b & 0\\ b & a & 0\\ 0 & 0 & 1\end{bmatrix} \tag{53}$$