Common Method Variance Techniques
Bradford R. Eichhorn, Cleveland State University, Cleveland, OH

ABSTRACT
Modern empirical research often utilizes survey tools such as Survey Monkey, Survey Gizmo, and Key Survey. Using a single survey respondent as the source for both the independent and dependent data in one instrument introduces the possibility of bias caused by using a single method of data collection. Additionally, the design of the survey instrument itself can cause raters to bias their responses. Common Method Variance (CMV) is the observation of such bias. The impact of this bias was beginning to be understood as long ago as the 1950s and is important because it introduces potentially significant errors in the measures.

This paper presents this common issue in empirical (survey) research, briefly lists techniques available to reduce this source of bias during the experiment’s design, and provides the analytical procedures to estimate the common method variance using three different techniques: Harman Single Factor, Common Latent Factor, and Common Marker Variable. We show how researchers can utilize PROC FACTOR and PROC CALIS in SAS® 9.2 to perform these three analyses.

INTRODUCTION
BACKGROUND
Bagozzi and Yi (1991, p. 426) define CMV as the “variance that is attributable to the measurement method rather than to the construct of interest”. Richardson et al (2009, p. 763) define CMV as “systematic error variance shared among variables measured with and introduced as a function of the same method and/or source.” There is little consensus regarding the veracity and magnitude of its impact. For example, Spector (2006, p. 230) says that CMV is an “urban legend” that is “an exaggeration and oversimplification of the true state of affairs.”

The reason for focusing attention on this subject is that the researcher’s conclusions are at risk since the conclusions regarding the model’s relationships may be erroneous (i.e., the error is too large for the relationships to be valid). For example, systematic correlations introduce an alternative explanation for the observed correlations between measures. Further, errors from the measurement instrument or method may have both random and systematic elements (Bagozzi & Yi, 1991). Campbell and Fiske (1959) note that errors introduced by methods and tools could contaminate analytical results. Cote and Buckley (1987) find that common method bias can vary considerably by discipline and the type of construct being investigated. Podsakoff et al (2003) summarize the literature showing evidence of CMV across disciplines and the extent of influence that CMV has between modeled relationships (this influence affects both the magnitude and direction of the relationships). The authors use Cote and Buckley’s (1987) estimate of average correlations to demonstrate that when common method bias is introduced, even when two constructs are perfectly correlated, the observed correlations can be as low as .52, and when two constructs have no correlation whatsoever, they can show a correlation as high as .23 because of random and systematic error (Podsakoff et al, 2003, p. 881). Not only can the strength of the bias vary causing the relationships between constructs to increase or decrease, but the direction of the effect may be impacted. These variances can increase both Type I and Type II errors.

Journals understand the need for implementing survey instruments. Ashkanasy (2008, p. 264) states that “authors need at a minimum to address potential threats to validity occasioned by common methods. While common methods issues are controversial in some respects (e.g., see Spector, 2006), they cannot be ignored.” Craighead et al (2011) state (in their review of IEEE research articles) that “CMV can distort observed relationships, thereby causing researchers to reach erroneous conclusions. As such, the unchecked presence of CMV undermines a study’s potential contributions to knowledge, which may be particularly problematic in survey research.” In response to this inherent measurement risk, journals are more frequently requiring researchers to demonstrate both proactive instrument design efforts and post-survey analyses of the amount of common method variance (Conway and Lance, 2010). They want authors to demonstrate that their survey methods have not introduced excessive variance that can alter the findings. The journals are sometimes requiring the use of multiple analytical tools since there is no consensus concerning which tools may be better than others.

This paper does not discuss the size or impact of CMV on research conclusions (a sample of such research includes Podsakoff et al, 2003; Richardson et al, 2009; and Williams et al, 2010), but instead presents three popular
implementations of post hoc statistical techniques to estimate such variance. This paper presents a research situation that is used throughout the subsequent analyses, followed by a review of the techniques, their SAS implementations, and ends with concluding remarks. However before continuing, we summarize potential sources of such variations, provide various practices to prevent or reduce the occurrence of common method variations, and present a review of latent variable modeling.

POTENTIAL SOURCES OF CMV
Reducing CMV in empirical studies begins with an understanding of its various potential sources. We refer the reader to Podsakoff et al (2003, pp. 881-885) and MacKenzie and Podsakoff (2012, pp. 544-545) for more complete discussions of potential sources. We summarize their findings here:

• The use of a common source or rater (one source that provides both independent and dependent variables) introduces a self-reporting bias. Their positive or negative perspectives of the research subject can influence each response to varying degrees.
• The survey instrument’s design, complexity, ambiguity and scale format can influence the rater’s responses.
• The item’s context (such as its position within the sequence of questions, its spatial relationship to other questions, and the number of questions) can affect the rater’s responses based on its stimulus to the rater.
• The survey’s measurement context can introduce covariation between measures. These characteristics include whether the independent and dependent variables are captured at the same point in time, in the same location or using the same medium.
• A rater’s motivation to answer accurately can be impacted based on the survey instrument’s characteristics such as the rater’s knowledge of the subject, their perceived ability to process and understand the subject, the length of the survey instrument and any inducements to respond, including “Don’t Know” or “Not Applicable” options if available.

PROACTIVE EXPERIMENT DESIGN
Although this paper focuses on post-hoc statistical detection of common method variance, there are a number of techniques available to the researcher that can reduce the chance and magnitude of such variations if applied prior to the experiment; i.e., during the design of the experiment itself. We summarize a selection of these techniques here and note that some are easily implemented while others may have meaningful benefits but potentially involve greater levels of time and money than may be available (see Chang et al, 2010, p. 179; Podsakoff et al, 2003, pp. 887-888; Craighead et al, 2011; and MacKenzie and Podsakoff, 2012, pp. 544-545):

• Obtain independent and dependent measures from separate sources (if possible). Although this can eliminate many sources of bias, it may be impractical if not impossible to implement.
• Separate the collection of independent and dependent variables by time, space, method or deception (concealing the primary purpose of the question to reduce the potential sensitivity of the question to the rater). These, too, significantly reduce or eliminate various sources of bias, but may introduce intervening factors thereby contaminating the results. Further, these techniques require additional time, effort and expense to implement.
• Protect the rater’s anonymity especially if the rater desires anonymity or there are corporate policies prohibiting survey involvement. This is simple to implement and can lead to responses that are more aligned with the research goal.
• Reduce the rater’s apprehension over their responses by communicating that this is a survey (there are no “right or wrong” answers) in an attempt to reduce the chance that they will edit their answers to give what they perceive as the best answers.
• Counterbalance or randomize the question order to disrupt potential interference between questions. Although this is simple to implement, it has the risk of disturbing the logical flow of the rater’s thought process which may degrade the quality of their responses.
• The researcher should seek to improve the scale items. The survey instrument should define terms, provide examples, maintain simplicity, and avoid complex syntax. Scale anchors should not be changed and reverse scoring should be limited because of the risk of reducing scale validity.
• Avoid “double-barreled” questions; i.e., one question should have only one subject. For example, one question should ask only about spinach and not ask their preference between spinach and turnips thereby introducing a second subject.

• Select respondents with sufficient experience to properly address each measure (i.e., properly identify the sample frame). Ensure that they can link key terms to concepts, have enough knowledge to retrieve for a suitable answer, and allow them to draw inferences to fill gaps when needed.

• A survey instrument should be pre-tested by a representative group of raters (practitioners and academics) to validate the instrument’s readability, clarity, length and its appropriateness for the sample frame. They should be prompted for feedback and suggested improvements and where appropriate, changes to the instrument should be implemented. Pre-testing also supports the instrument’s content validity.

• Include specific marker variable measures if the literature supports theoretically uncorrelated measures.

LATENT VARIABLE MODELING

Since much research can be completed without the use of latent variables, we provide a brief explanation of manifest and latent variables for readers unfamiliar with the terms and concepts. Latent variables are essentially unobservable; they must be inferred through the observation of other measures. The observable measures are called manifest variables. Shah and Goldstein (2006, p. 149) state that latent variables can be exogenous (independent) or endogenous (dependent) and are typically represented by multiple directly measurable variables.

Models that use latent variables may be seen in many disciplines including econometrics, artificial intelligence, bioinformatics, physics and management. One example of such a model would be the measurement of intelligence. Since there is no practical observation of “intelligence”, we infer this characteristic by measuring scores on various tests examining problem solving, abstract thinking, reasoning, planning skills, and more. Another example from finance could be the domestic performance of a company (the latent variable) as measured by various manifest variables such as customer orders, production, inventory, and the number of full- and part-time employees.

The number of manifest variables used to identify a latent variable has been studied since that affects the latent variable’s validity. Kline (2005, p. 314) suggests that although two manifest variables might be sufficient, three or more are better for the reduction of specification errors. Hatcher (1994, p. 260) recommends the use of at least three manifest variables per latent variable. Latent variables with less than five manifest variables can exhibit problematic results if they occur in models with small sample sizes (Johnson and Creech, 1983). However, the authors state that the “distortions [are] not of sufficient magnitude to strongly bias the estimates of important variables” (ibid, p. 406). Having a sufficient number of manifest variables is important because a confirmatory factor analysis may find that one or more of the manifest variables do not associate with their intended latent variable; then those manifest variables would be removed from the model. An exploratory factor analysis may be called for if the research is determining new or modified latent variables or their observable measures; again this may determine that the measures do not associate with the hypothesized factors. The recommendations by Kline (2005), Hatcher (1994) and Johnson and Creech (1983) refer to the number of manifest variables remaining after removing those that did not load as hypothesized.

RESEARCH SCENARIO

For this paper, we introduce a simple generic model containing two latent factors measured by nine manifest variables (Figure 1). We note that common method bias only affects observed (manifest) variables. If there are one or more second-order latent variables, the analyses should be performed using only the manifest variables. Also, biases are observed because of societal, instrumental or temporal factors. Objective manifest variables (machine time, bank balances, laboratory measurements, etc.) are not subject to method bias and should not be included in common method variance analyses.
THREE ANALYTICAL TECHNIQUES: AN INTRODUCTION

The goal of testing for common method variance is to determine to what degree any such biases exist. Analytical techniques estimate the degree to which the data may be influenced by biases caused by the survey method or tool. We describe three frequently used techniques to estimate common method variance.

HARMAN SINGLE FACTOR

This first technique (Harman, 1960) uses exploratory factor analysis where all variables are loaded onto a single factor (F3 in Figure 2) and constrained so that there is no rotation (Podsakoff et al, 2003). This new factor is typically not in the researcher’s model; it is introduced solely for this analysis and then discarded. If the newly introduced common latent factor explains more than 50% of the variance, then common method bias may be present.

The Harman single factor technique has the benefit of simplicity. However, there are multiple weaknesses with this method.

- It doesn’t statistically control for this type of variance.
- There are no specific guidelines on the amount of variation explained by this factor to determine unequivocally the existence of this variance. The customary heuristic is to set the threshold to 50%.
- The method is sensitive to the number of variables involved. Large models have a greater chance for multiple common method factors to exist. As the number of variables increases, this technique becomes less conservative.
- The sample may be subject to multiple sources of bias but this technique assumes a single source which potentially misrepresents the actual bias(es).
Figure 2: Harman Single Factor Technique

This second technique introduces a new latent variable in such a way that all manifest variables are related to it, those paths are constrained to be equal and the variance of the common factor is constrained to be 1 (See Figure 3). This is similar to the Harman Single Factor technique where all manifest variables are related to a single factor; however the research model’s latent factors and their relationships are kept in this analysis. The common variance is estimated as the square of the common factor of each path before standardization. The common heuristic is to set the threshold to 50%.

This technique allows for measurement error, focuses on the measures themselves, and doesn’t require the researcher to identify and measure the specific factor responsible for common method effects. However, it assumes no interaction with the constructs and doesn’t allow the researcher to insert any known or suspected cause(s) of bias. Therefore, the method factor (F3) may actually represent multiple biases, similar to the Harman Single Factor technique. This technique aspires to allow small models to be identified by setting all paths to F3 equal, but that undermines the advantage of allowing each measure’s loading to vary (Kline, 2005, p. 105, defines structured equations as being identified “if it is theoretically possible to derive a unique estimate of each parameter”).

Figure 3: Common Latent Variable

This third technique (Figure 4) allows the researcher to include measures presumed to influence the cause of the bias itself. The survey instrument would ask for measures of these influences that are loaded onto the new method factor (F4) with all manifest variables (V1-V12) associated with a common method factor (F3). The loading of the common method manifest variables are forced to be equal. Lindell & Whitney (2001) note that there is some dispute
concerning whether these loading parameters should or should not be set equal to each other since one wouldn’t expect their values to be equal. However, the research question isn’t whether they are equal, but whether they are significant enough to alter the correlations of the research model. The authors note that for self-reported values of dependent variables, it is plausible that these loadings may be equal. The common variance in this technique is the square of the common factor of each path before standardization. Again, the common heuristic is to set the threshold to 50%.

There are a number of advantages to this technique. First, it allows measurement error in the method factor (F4) to be estimated. Second, the effects of biases are measured directly rather than being inferred from the model’s measures. Third, the impacts of each measure in the method factor are not constrained to be equal. However, this requires that the researcher know the most important sources of method bias and construct an appropriate collection instrument. This can be a significant weakness since the sources of bias may not be understood at all let alone with sufficient detail to model. Even if understood, some may not be measurable due to psychological factors or those implied within the survey instrument itself.

We note that if a viable marker variable and its measures were not collected as part of the data collection process, the Common Latent Factor technique may be a better alternative. Richardson et al (2009) studied over 62,000 simulations of common method variance using various scenarios. They define “ideal markers” as variables with no expected theoretical relationship with substantive variables” (ibid, p. 768). They further state that this technique with the use of an ideal marker is a reasonable tool to detect the presence of common method variance.

If the research model was sufficiently large, the researcher could use multiple uncorrelated measures from the study itself. Lindell & Whitney (2001) suggest using variables that have very low correlations between manifest variables as measures for the latent method variable.

Figure 4: Common Marker Variable

THREE ANALYTICAL TECHNIQUES: SAS IMPLEMENTATIONS

We now provide the SAS implementations to these analytical techniques and include only the significant portions of the SAS output to demonstrate each technique. The complete dataset and code can be requested by contacting the author (see Contact Information at the end of this paper).

HARMAN SINGLE FACTOR

This analysis employs PROC FACTOR to calculate the eigenvalues for an analysis of all manifest variables being loaded onto a single factor. Note that principal component analysis is specified for a single factor without rotation.

```sas
TITLE "Harman";
PROC FACTOR DATA = CMV.Data
   METHOD = PRINCIPAL
```
The sample data contains 195 valid data points and generated the following list of eigenvalues:

<table>
<thead>
<tr>
<th>Eigenvalue</th>
<th>Difference</th>
<th>Proportion</th>
<th>Cumulative</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4.87440840</td>
<td>0.5416</td>
<td>0.5416</td>
</tr>
<tr>
<td>2</td>
<td>1.20135706</td>
<td>0.1335</td>
<td>0.6751</td>
</tr>
<tr>
<td>3</td>
<td>0.74963657</td>
<td>0.0833</td>
<td>0.7584</td>
</tr>
<tr>
<td>4</td>
<td>0.64104606</td>
<td>0.0712</td>
<td>0.8296</td>
</tr>
<tr>
<td>5</td>
<td>0.43355117</td>
<td>0.0482</td>
<td>0.8778</td>
</tr>
</tbody>
</table>

From this output, we see from the first row that the Harman Single Factor technique estimates the common method variance to be 54.2% which exceeds the commonly accepted threshold of 50%; this suggests that common method bias may be a problem with this dataset.

**COMMON LATENT FACTOR**

This analysis uses PROC CALIS to estimate the common method variance while simultaneously calculating the research model's covariances. Note that F3 is introduced with a single loading parameter of “CLF” which causes all of those loadings to be equal, the variance of F3 is set to 1 as specified, and F3 is not covaried with either F1 or F2 (see Figure 3).

```
TITLE "Common Latent Factor";
PROC CALIS DATA = CMV.Data PLOTS = RESIDUALS CORR RESIDUAL MAXITER = 40000 MODIFICATION;
LINEQS V1 = LV1F1 F1 + CLF F3 + E1,
       V2 = LV2F1 F1 + CLF F3 + E2,
       V3 = LV3F1 F1 + CLF F3 + E3,
       V4 = LV4F1 F1 + CLF F3 + E4,
       V5 = LV5F1 F1 + CLF F3 + E5,
       V6 = LV6F2 F2 + CLF F3 + E6,
       V7 = LV7F2 F2 + CLF F3 + E7,
       V8 = LV8F2 F2 + CLF F3 + E8,
       V9 = LV9F2 F2 + CLF F3 + E9;
VARIANCE F1 = 1,
          F2 = 1,
          F3 = 1,
          E1 = VARE1,
          E2 = VARE2,
          E3 = VARE3,
          E4 = VARE4,
          E5 = VARE5,
          E6 = VARE6,
          E7 = VARE7,
          E8 = VARE8,
          E9 = VARE9;
COV F1 F2 = CF1F2,
     F1 F3 = 0,
     F2 F3 = 0;
VAR V1 V2 V3 V4 V5 V6 V7 V8 V9;
RUN;
```
The following partial output is from the linear equations output before standardization (not all variables are shown):

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>T Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>V1</td>
<td>-0.5709<em>F1  + 0.6470</em>F3 + 1.0000 E1</td>
<td>0.0691 LV1F1 0.0420 CLF</td>
<td>-8.2635</td>
</tr>
<tr>
<td>V2</td>
<td>-0.4391<em>F1  + 0.6470</em>F3 + 1.0000 E2</td>
<td>0.0705 LV2F1 0.0420 CLF</td>
<td>-6.2288</td>
</tr>
<tr>
<td>V3</td>
<td>-0.3532<em>F1  + 0.6470</em>F3 + 1.0000 E3</td>
<td>0.0815 LV3F1 0.0420 CLF</td>
<td>-4.3363</td>
</tr>
<tr>
<td>V4</td>
<td>-0.3881<em>F1  + 0.6470</em>F3 + 1.0000 E4</td>
<td>0.0758 LV4F1 0.0420 CLF</td>
<td>-5.1207</td>
</tr>
</tbody>
</table>

Note that the CLF value = 0.6470 for all variables shown and their t-value indicates significance. The common method variance is the square of that value, $0.6470^2 = 0.4186$. Therefore, the Common Latent Factor technique suggests that there is no significant common method bias in this data since the calculated variance (41.9%) is below the threshold of 50%.

**COMMON MARKER VARIABLE**

This analysis also uses PROC CALIS but with an additional marker variable (F4) added into the model. For this example, F4 is measured using three manifest variables that are believed to be uncorrelated with any other variables in the study. We retain the use of F3 and its single loading parameter of “CLF”. Note that F3 is not allowed to covary with F4 but F4 does covary with both F1 and F2 (see Figure 4).
The following partial output is from the linear equations output before standardization (not all variables are shown):

\[
\begin{align*}
V1 &= -0.7509 \cdot F1 + 0.3750 \cdot F3 + 1.0000 \cdot E1 \\
\text{Std Err} &= 0.0748 \\
\text{t Value} &= -10.0402 \\
V2 &= -0.6771 \cdot F1 + 0.3750 \cdot F3 + 1.0000 \cdot E2 \\
\text{Std Err} &= 0.0760 \\
\text{t Value} &= -8.9149 \\
V3 &= -0.5215 \cdot F1 + 0.3750 \cdot F3 + 1.0000 \cdot E3 \\
\text{Std Err} &= 0.0800 \\
\text{t Value} &= -6.5217 \\
V4 &= -0.5462 \cdot F1 + 0.3750 \cdot F3 + 1.0000 \cdot E4 \\
\text{Std Err} &= 0.0766 \\
\text{t Value} &= -7.1305
\end{align*}
\]

Note that the CLF value is 0.3750 for all variables shown and its t-value indicates significance. The common method variance is the square of that value, \(0.3750^2 = 0.1406\). Therefore, the Common Marker Variable technique suggests that there is no significant common method bias in this data since the calculated variance (14.1%) is below 50%.

CONCLUSION

We briefly reviewed what Common Method Variance is and how it can be subtly introduced as part of an empirical study. We also provided a list of pro-active steps to reduce or eliminate such variances. The body of this paper presented three alternative methods for detecting and measuring the presence and significance of CMV.

Since “an ounce of prevention is worth a pound of cure” (The Quotable Franklin), we agree with Lindell & Whitney (2001) who recommend designing empirical experiments with proactive steps to reduce such variations. However, we must also be able to demonstrate the amount of common method variance in our studies and for that purpose, this paper presented three conventional tools available in SAS to perform such analyses.

Although the Harman Single Factor technique suggested that there was common method bias in our sample data, neither the Common Latent Factor nor the Common Marker Variable techniques suggested such a bias. Podsakoff et al (2003) and MacKenzie & Podsakoff (2012) state that the Harman Single Factor technique is not adequate so its findings may be discarded in light of better techniques. Lindell & Whitney (2001) and Kemery & Dunlap (1986) point out that although the Harman Single Factor method can be used, its tendency to remove construct variance along with common method variance limits its use to a last resort.

Lindell & Whitney (2001) recommend the use of a well-designed Common Marker Variable (one where the method factor’s measures are injected between topics within the survey instrument). This approach will also improve discriminant validity if the method factor is supported by a multi-item scale and is theoretically unrelated to at least one research construct. Williams et al (2010) review published studies employing various forms of marker variables. They state that “adding a marker latent variable and its indicators to [experimental designs] will always yield stronger analyses and is likely to be feasible in most circumstances.” (Ibid, p. 505). They note that it is possible to use multiple method variables when the theory supports such distinguishing factors. Regarding the topic of selecting marker...
variables and their indicators, they propose that researchers have specific biases and substantive variables in mind and make a theoretical link between the marker variable and one or more of the biases they suspect.

This paper reviewed three basic model types. Podsakoff et al (2003), Williams et al (2010), Richardson et al (2009), and Skrondal and Rabe-Hesketh (2007) study common method variance using multiple models; many of the additional models are hybrid combinations of these basic models. Researchers should become familiar with the various available analytical models and their techniques to apply the most appropriate techniques to suit their research study and its unique characteristics.

In conclusion and with the goal of helping the researcher, a well-designed empirical study should consider the potential for common method bias, proactively address the issue during the experiment’s design, and include explicit measures to reduce potential biases as the best research option given the current state of knowledge regarding common method variance. The presentation of research results from multiple CMV analytical techniques is also recommended since there are many potential sources of bias that can affect a research study.

REFERENCES


ACKNOWLEDGMENTS

I would like to thank Dr. Walter Rom at Cleveland State University for motivating me to research this topic in depth and then asking me to present my research on these techniques to his DBA students in Multivariate Data Analysis. Additionally, comments and suggestions from both Doug Thompson and Michael Wilson were instrumental in improving this work well past its original content.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. The complete dataset and code can be requested by contacting the author at:

Bradford R. Eichhorn, PMP
Visiting Assistant Professor, Department of Operations & Supply Chain Management
Monte Ahuja College of Business Administration
BU 534
Cleveland State University
Cleveland, OH 44114
Phone: 216 687 3657
b.eichhorn@csuohio.edu

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.