# WEBSITE APPENDIX A

## Personality, Abilities and Motivations

Table 1. Individual differences widely studied in psychology

<table>
<thead>
<tr>
<th>Individual Difference</th>
<th>Definition (APA Dictionary definition in quotes)</th>
<th>Example Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Personality Traits: How individuals typically act, think, and feel</td>
<td>“A model of the primary dimensions of individual differences in personality. The dimensions are usually labeled extraversion, neuroticism, agreeableness, conscientiousness, and openness to experience, thought he labels vary somewhat among researchers.”</td>
<td>NEO-PI-R (Costa &amp; McCrae, 1992); Big Five Inventory (John &amp; Srivastava)</td>
</tr>
<tr>
<td>Big Five personality model</td>
<td>“A dimension of the Big Five personality model that refers to individual differences in the tendency to be open to new aesthetic, cultural, or intellectual experiences.” Openness correlates with IQ ($r$ about .3). Also called Intellect, this dimension includes facets such as open-mindedness, creativity, appreciation of arts and music.</td>
<td>Typical Intellectual Engagement (TIE, Ackerman); Need for Cognition (Cacioppo); Openness domain of any Big Five questionnaire</td>
</tr>
<tr>
<td>Big Five Openness to Experience</td>
<td>“The tendency to be organized, responsible, and hardworking, construed as one end of a dimension of individual differences (conscientiousness vs. lack of direction) in the Big Five personality model.” Conscientiousness predicts work and health outcomes and includes facets such as dependability, orderliness, perseverance, and need for achievement.</td>
<td>Conscientiousness domain of any Big Five questionnaire</td>
</tr>
<tr>
<td>Big Five Conscientiousness</td>
<td>“One of the dimensions of the...Big Five personality model characterized by a chronic level of emotional instability and proneness to psychological distress.” This dimension is often termed emotional stability, which is the opposite of neuroticism, and includes facets such as hostility, depression, and anxiety.</td>
<td>Neuroticism domain of any Big Five questionnaire. Negative affect measures could also be used but the latter often emphasize temporary affect rather than dispositional affect.</td>
</tr>
<tr>
<td>Big Five Neuroticism/Emotional Stability</td>
<td>“The tendency to act in a cooperative, unselfish manner, construed as one end of a dimension of individual differences (agreeableness vs. disagreeableness) in the Big Five personality model.” This dimension includes facets such as trust and compliance.</td>
<td>Agreeableness domain of any Big Five questionnaire.</td>
</tr>
<tr>
<td>Big Five Agreeableness</td>
<td>“An orientation of one’s interests and energies toward the outer world of people and things rather than the inner world of subjective experience. Extraversion is a broad personality trait and, like introversion, exists</td>
<td>Extraversion domain of any Big Five questionnaire.</td>
</tr>
<tr>
<td>Trait</td>
<td>Description</td>
<td>Measures</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Social vitality</td>
<td>A subdimension of Big Five Extraversion proposed by Roberts (2006) that includes facets such as sociability, positive effect, and gregariousness.</td>
<td>CPI Sociability Scale; NEO-PI-R Gregariousness and Activity Scales</td>
</tr>
<tr>
<td>Social dominance</td>
<td>A subdimension of Big Five Extraversion that includes facets such as dominance, independent, and self-confidence, especially in social settings.</td>
<td>NEO-PI-R Assertiveness scale; 16PF Dominance Scale</td>
</tr>
<tr>
<td>Need for Achievement/Ambition</td>
<td>“A strong desire to accomplish goals and attain a high standard of performance and personal fulfillment. People with high need for achievement undertake tasks in which there is a reasonable probability of success and avoid tasks that are too easy or too difficult”</td>
<td>Projective measures (unreliable), Conscientiousness subscales</td>
</tr>
<tr>
<td>Delay of gratification/Impulsivity/Self-Control/Time Preference</td>
<td>“Forgoing immediate reward in order to obtain a larger or more desirable reward in the future” A facet of either Neuroticism or Conscientiousness. When faced with a choice between immediate temptation and superior, deferred gratification, the ability to control behavior, attention, and thoughts in order to attain superior, deferred reward.</td>
<td>Marshmallow task (Mischel); UPPS Impulsive Behavior Subscale (Whiteside &amp; Lynam); Self-Control Scale (Baumeister); Time Perspective Inventory (Zimbardo)</td>
</tr>
<tr>
<td>Sensation Seeking</td>
<td>“The tendency to search out and engage in thrilling activities as a method of increasing stimulation and arousal. It takes the form of engaging in highly stimulating activities accompanied by a perception of danger.” A facet of either Conscientiousness or Extraversion. Attraction to novel, intense experiences and willingness to take risks for them.</td>
<td>Sensation-seeking Scale</td>
</tr>
<tr>
<td>Perceived Self-efficacy/locus of control/optimism</td>
<td>Perceived Self Efficacy: “Subjective perception of own capability of performance or ability to attain results”. Locus of Control: “perception of how much control individuals have over conditions of their lives”. Possibly a facet of Neuroticism. The belief that one has control over outcomes. The opposite of helplessness.</td>
<td>Generalized self-efficacy scales⁴¹, Rotter Locus of Control Scale, Attributional Style Questionnaire</td>
</tr>
</tbody>
</table>

⁴¹ Bandura does not endorse any trait-level scales
<table>
<thead>
<tr>
<th><strong>Regulatory Focus</strong></th>
<th>Orientation toward either promotion of positive outcomes or prevention of negative outcomes.</th>
<th><strong>Regulatory Focus Questionnaire</strong> (Higgins et al., 2001)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Affect, Well-Being</strong></td>
<td>Positive affect is the internal feeling state that occurs when a goal has been attained, a source if threat has been avoided or the individual is satisfied with the present state of affairs”. Positive affect includes emotions such as joy, contentment, and pride. Negative affect, which is only moderately inversely correlated with positive affect, includes fear, anxiety, sadness, etc. Life satisfaction is a cognitive, not affective, appraisal of the quality of one’s life. Well-being is characterized by the presence of positive affect, the absence of negative affect, and positive life satisfaction.</td>
<td>PANAS (Watson, Clark, &amp; Tellegen) for positive and negative affect; SWLS (Diener) for life satisfaction</td>
</tr>
<tr>
<td><strong>Self-esteem</strong></td>
<td>“The degree to which the qualities and characteristics contained in one’s self concept are perceived to be positive” One's estimation of one's own self-worth. A construct that enjoyed tremendous popularity in the 1970s but since has been considered epiphenomenal not causal. A minority of psychologists consider positive and negative evaluations of the self to be the sixth and seventh factors of personality. Possibly grouped with self-efficacy, etc.</td>
<td>Rosenberg Self-Esteem Scale</td>
</tr>
<tr>
<td><strong>Temperament (childhood)</strong></td>
<td>“Basic foundation of personality, usually assumed to be biologically determined and present early in life[...] Includes characteristics such as energy level, emotional responsiveness, response tempo and willingness to explore” Precursors to personality traits, temperament variables are patterns in behavior or affect that appear early in life that are assumed to have a neurobiological basis. There are fewer temperament variables than personality traits because these individual differences are less salient and stable in children than in adults.</td>
<td>Children's Behavior Questionnaire</td>
</tr>
<tr>
<td><strong>Psychopathology</strong></td>
<td>“Patterns of behavior or thought processes that are abnormal or maladaptive”. A broad category comprising dysfunctional patterns of thought, feeling, or behavior. Most disorders are included in the DSM-IV manual. Axis I disorders (e.g., depression) are more intense and episodic/discreet, whereas Axis II disorders (i.e., personality disorders) are more tonic and enduring.</td>
<td>Beck Depression Inventory; Beck Anxiety Inventory; MMPI (omnibus measure); Child Behavior Checklist (Achenbach)</td>
</tr>
<tr>
<td><strong>Myers-Briggs Type Indicator (MBTI)</strong></td>
<td>“A personality test designed to classify individuals according to their expressed choices between contrasting alternatives in certain categories of traits. The categories, based on Jungian typology, are extraversion-extraversion, Sensing-Intuition, Thinking-Feeling, and Judging-Perceiving...The test has little credibility among research psychologists but is widely used in educational counseling and human resource management...”</td>
<td></td>
</tr>
<tr>
<td><strong>Type A/Type B personality</strong></td>
<td>Type A personality is “a personality pattern characterized by chronic competitiveness, high levels of achievement motivation, and hostility.” Type B personality is “a personality pattern characterized by low levels of competitiveness and frustration and a relaxed, easy going approach.”</td>
<td></td>
</tr>
<tr>
<td><strong>Intelligence, General Mental Ability, “g”, IQ</strong></td>
<td>Intelligence and general mental ability both refer to “the ability to derive information, learn from experience, adapt to the environment, understand and correctly utilize thought and reason.” “g” (“general factor”) refers to the first factor extracted from a factor analysis of cognitive tasks, which many researchers consider to represent general (vs. specific) intelligence.”It represents individuals’ abilities to perceive relationships and to derive conclusions from them. It is the basic ability that underlies the performance of different intellectual tasks”</td>
<td></td>
</tr>
<tr>
<td><strong>Specific mental abilities</strong></td>
<td>“Abilities as measured by tests of an individual in areas of spatial visualization, perceptual need, number facility, verbal comprehension, word fluency, memory, inductive reasoning and so forth” An umbrella category for lower-level mental abilities, including math, verbal, and spatial abilities, as well as even more specific mental capacities.”</td>
<td></td>
</tr>
<tr>
<td><strong>Creativity</strong></td>
<td>“Ability to produce original work, theories, techniques or thoughts […] Related with imagination, expressiveness, originality.” Ability to generate novel ideas and behaviors that solve problems.</td>
<td></td>
</tr>
</tbody>
</table>

42 Cattell considered fluid (capacity to learn) and crystallized intelligence (knowledge) to be second-order aspects of intelligence.
<table>
<thead>
<tr>
<th>Executive Function</th>
<th>“Higher level cognitive processes that organize and order behavior, including logic and reasoning, abstract thinking, problem solving, planning and carrying out and terminating goal directed behavior” Broad set of higher-level cognitive capacities attributed to the prefrontal cortex, often involving the coordination and management of lower-level processes.</th>
<th>Innumerable neuropsychology tasks (e.g., go/no-go, Stroop, Continuous Performance Task)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognitive reflection</td>
<td>A specific mental ability. The tendency to reflect before taking an intuitive answer as correct.</td>
<td>Cognitive Reflection Test (Frederick, 2005)</td>
</tr>
<tr>
<td>Emotional Intelligence</td>
<td>“Ability to process emotional information and use it in reasoning and other cognitive activities. According to Mayer and Slovey 1997 model it comprises four abilities: to perceive and appraise emotions accurately, to access and evoke emotions when they facilitate cognition, to comprehend emotional language and make use of emotional information, and to regulate one’s own and others’ emotions to promote growth and well-being” Ability to perceive emotion, to integrate it in thought, to understand it and to manage it (Mayer)</td>
<td>MSCEIT (Mayer &amp; Salovey) but really there are no good measures</td>
</tr>
<tr>
<td>Motivation – What individuals want to do, feel, or think</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Values</td>
<td>“A moral, social or aesthetic principle accepted by an individual (or society) as a guide to what is good, desirable or important.” What individuals feel is important in a moral sense.</td>
<td>Values in Action Inventory of Strengths</td>
</tr>
<tr>
<td>Interests</td>
<td>“Attitude characterized by a need to give selective attention to something that is significant to the individual”. What spontaneously attracts and holds one’s attention, and is considered pleasant.</td>
<td>Self-Directed Search, Strong Interest Inventory</td>
</tr>
<tr>
<td>Goals, Needs, Motives</td>
<td>Goal: “The end state toward which a human is striving” Motive:”physiological or psychological state of arousal that directs an organism’s energies toward a goal”. What individuals aim to achieve or experience. Examples include need for power, need for affiliation, and need for achievement. Note that many consider need for achievement as an aspect of personality.</td>
<td>Thematic Apperception Test (TAT), need for achievement questionnaires</td>
</tr>
</tbody>
</table>
Problem similar to the Raven's Progressive Matrices test items

Note: The bottom right entry of this 3x3 matrix of figures is missing and must be selected from among 8 alternatives. Looking across the rows and down the columns, the test taker attempts to determine the underlying pattern and then pick the appropriate missing piece. The correct answer to this problem is 5. Figure taken from Carpenter, Just, and Shell (1990), used with permission of the publisher, copyright American Psychological Association.
Measurement Error Can Create the Illusion of Multiple Factors when Only One is Operative

If some outcome $Y_j$ is predicted by $f_k$, and there are multiple mismeasured proxies for $f_k$, those proxies will be predictive for $Y_j$ even though only one factor generates the outcome. Thus, if the following relationship holds between $Y_j$ and $f_k$, where $U_j$ is statistically independent of $f_k$,

$$Y_j = \alpha + \beta f_k + U_j,$$

and we use $N$ $Q$-adjusted proxies for $f_k$

$$\tilde{M}_k^n = M_k^n - \mu_k^n(Q) = \lambda_k^n f_k + \varepsilon_k^n$$

where $\varepsilon_k^n$ is independent of $f_k$, has mean zero and variance $\sigma_{\varepsilon_k^n}^2$, where we allow the means of the measures to depend on $Q$ ($\mu_k^n(Q)$), we obtain an equation

$$Y_j = \tilde{\alpha} + \sum_{n=1}^{N} \gamma_n \tilde{M}_k^n + \tilde{U}_j,$$

where $\tilde{U}_j = \beta f_k + U_j$ and where the true values of $\gamma_n$ are all zero, because the test scores do not determine $Y_j$ but instead $f_k$ determines it unless $\beta = 0$. Straightforward calculations show that under general conditions, if one arrays the $\gamma_n$ in a vector of length $N$, $\gamma^N$, and the $Q$-adjusted test scores in a vector of length $N$, denoted $\tilde{M}$, the OLS estimator

$$\hat{\gamma}^N = \left(\text{Cov}(\tilde{M},\tilde{M})\right)^{-1} \text{Cov}(\tilde{M},Y)$$

converges in large samples to
\[
\text{plim } \gamma^N = \beta \sigma_{s_i}^2 \begin{pmatrix}
\sigma_{s_i}^2 + \left( \lambda_{k}^1 \right)^2 \sigma_{f_k}^2 & \lambda_{k}^1 \lambda_{k}^2 \sigma_{f_k}^2 & \cdots & \lambda_{k}^1 \lambda_{k}^N \sigma_{f_k}^2 \\
\lambda_{k}^1 \lambda_{k}^2 \sigma_{f_k}^2 & \sigma_{s_i}^2 + \left( \lambda_{k}^2 \right)^2 \sigma_{f_k}^2 & \cdots & \vdots \\
\vdots & \cdots & \cdots & \vdots \\
\lambda_{k}^1 \lambda_{k}^N \sigma_{f_k}^2 & \cdots & \sigma_{s_i}^2 + \left( \lambda_{k}^N \right)^2 \sigma_{f_k}^2 \\
\end{pmatrix}^{-1} \begin{pmatrix}
\lambda_{k}^1 \\
\vdots \\
\lambda_{k}^N \\
\end{pmatrix}.
\]

Assuming \( \sigma_{s_i}^2 > 0, \beta \neq 0 \) and \( \sigma_{s_i}^2 > 0 \), for all \( n = 1, \ldots, N \), any error-ridden predictor of \( f_k \) will be statistically significant in a large enough sample. Using a purely predictive criterion to determine personality traits produces a proliferation of “significant” predictors for outcome \( Y_j \), as is found in the psychological studies that we survey in the text. Cunha and Heckman (this issue) show that estimated measurement errors in both cognitive and noncognitive tests are important so that the problem of proxy proliferation using a predictive criterion is serious. Under such a criterion, many “significant” predictors of an outcome can be found that are all proxies for a single latent construct.

If, in addition to these considerations, the measures fail discriminant validity, the predictors for one outcome may proxy both \( f_k \) and \( f_{k'} \) \( (k \neq k') \) if \( f_k \) and \( f_{k'} \) are correlated. We present evidence in the text, Section III, that IQ tests proxy both cognitive and personality factors. A purely predictive criterion fails to distinguish predictors for clusters associated with the \( f_k \) from predictors for clusters proxying the \( f_{k'} \). Thus, items in one cluster can be predictive of outcomes more properly allocated to another cluster.

**Accounting for Reverse Causality**

A test score may predict an outcome because the outcome affects the test score.

Hansen, Heckman, and Mullen (2004) analyze a model in which, in the previous representation, the outcome being related to cluster (factor) \( k \), say \( Y_j \), is an element of \( Q \)
and determines $\mu_k^n(Q)$ and $\lambda_k^n(Q)$. In addition, they allow (factor) $f_k$ to be a determinant of $Y_j$. They establish conditions under which it is possible to identify causal effects of $f_k$ on $Y_j$ when the proxies for $f_k$ suffer from the problem of reverse causality because the $Q$ in $\lambda_k^n(Q)$ and $\mu_k^n(Q)$ may include $Y_j$ among its components. They establish that tests of cognitive ability (e.g., AFQT) are substantially affected by schooling levels at the date the test is taken.

To understand how their method works at an intuitive level, consider the effect of schooling on measured test scores. Schooling attainment likely depends on true or “latent” ability, $f_k$. At the same time, the measured test score depends on schooling attained so it affects $\mu_k^n(Q)$ or $\lambda_k^n(Q)$ or both. Hansen, Heckman, and Mullen (2004) assume access to longitudinal data that randomly sample the population. Included in the sample are adolescents. At the time the adolescents are given the test, persons who eventually attain the same schooling level are at different grade levels at the date of the test because the longitudinal sample includes people of different ages and schooling levels. From the longitudinal data we can determine final schooling levels. Final schooling levels are assumed to depend on latent ability $f_k$. Conditional on the final schooling level attained, schooling levels at the date of the test are random with respect to $f_k$ because the sampling rule is random across ages at a point in time. One can identify the causal effect of schooling on test scores from the effect of variation in the years of education attained at the date of test on test scores for persons who attain the same final schooling level. See Hansen, Heckman, and Mullen (2004) for additional details on this method and alternative identification strategies. The basic idea of their procedure is to
model the dependence between $Q$ and $f_k$ and to solve the problem of reverse causality using this model. Heckman, Stixrud and Urzua (2006) develop this method further. Cunha and Heckman (2007) and Cunha, Heckman and Schennach (2006) extend this procedure to allow $f_k$ to evolve over time through investment and experience.
a) Personality

a1. “BIG FIVE”:

The literature on personality psychology widely adopts “the big five” as a taxonomy for describing one’s personality traits. The idea is that there are five big dimensions over which personality can be studied. These dimensions are openness to experience, conscientiousness, neuroticism, agreeableness and extraversion. Each dimension is then subdivided in lower level traits called “facets”. While there is a relatively broad consensus over the big five, there is still a wide disagreement over which facets correspond to each dimension. In the present scheme, we propose one of the possible lower order structure.

Measures for the big five:

  http://www.uoregon.edu/~sanjay/pubs/bigfive.pdf

For a general introductory literature on the “big five”:


For a discussion of facets:


For further readings on the links between big five and smoking, crime, scholastic achievement and labor market outcomes:

1) on smoking and alcohol use:

2) on crime

3) on scholastic achievement
- Borghans L., Meijers H. and Ter Weel B., (2006), “The role of non cognitive skills in explaining non cognitive test scores” Maastricht University working paper, Maastricht, the Netherlands.
- Duckworth A., Seligman M.(2005),” Self Discipline outdoes IQ in predicting Academic Performance in Adolescents” Psychological Science 16, 934-944; on self control and scholastic achievement
- Nolfle and Robins, in press. I can’t find it!

4) on labor market performance

**DOMAINS OF THE BIG FIVE:**

1. **Openness to experience**

“A dimension of the Big Five personality model that refers to individual differences in the tendency to be open to new aesthetic, cultural, or intellectual experiences.” (APA Dictionary)

It is the Big Five domain that correlates most with IQ (r about .3). It is also called Intellect. Includes facets such as open-mindedness, creativity, appreciation of arts and music.

For its correlation with openness to experience, authors include into this category also J. Cacioppo’s **need for cognition** concept: an individual’s tendency to engage in and enjoy effortful cognitive endeavours. For a **measure of need for cognition** see the 18 items questionnaire at Cacioppo, Kao, Petty (1984), [http://www.leaonline.com/doi/pdfplus/10.1207/s15327752jpa4803_13](http://www.leaonline.com/doi/pdfplus/10.1207/s15327752jpa4803_13). Another measure related to openness to experience is the **TIE**, Typical Intellectual Engagement (see Ackerlof and Goff, 1992, 1994)

**For readings on Openness to Experience:**

2. **Conscientiousness**

“The tendency to be organized, responsible, and hardworking, construed as one end of a dimension of individual differences (conscientiousness vs. lack of direction) in the Big Five personality model.” (APA Dictionary)

Conscientiousness is the Big Five domain that best predicts work and health outcomes.
Most of the work on conscientiousness is done by Brent Roberts. His website link below is a good source for further readings.

Two well studied facets empirically related to conscientiousness are need for achievement and delay for gratification.

- Need for achievement is defined as “A strong desire to accomplish goals and attain a high standard of performance and personal fulfillment. People with high need for achievement undertake tasks in which there is a reasonable probability of success and avoid tasks that are too easy or too difficult”

- While Delay for gratification is the ability of foregoing immediate reward in order to obtain a larger or more desirable reward in the future” (APA Dictionary)

The “Marshmallow test” (Mischel 1989) was the first study of the relationship between delayed gratifications and outcomes later in life. See literature below. There is no single measure for delay of gratification.

As for need for achievement, it is usually measured by McClelland N-Ach Tematic Apperception Test (TAT), a projective test that however has come under serious critique from mainstream psychology.

For readings on Conscientiousness
- Brent Roberts papers on conscientiousness are available at his website, http://www.psych.uiuc.edu/~broberts/Brent%20W%20Roberts%20Research%20Interests.htm
- Need for Achievement:
- Delay of Gratification:
  - Funder, Block (1989) “The role of Ego-Control, Ego-Resiliency and IQ Delay of Gratification in Adolescence”. Journal of Personality and Social Psychology, 57(6), 1041–1050. A delay gratification experiment with 14 years old adolescent, evidence that delay behaviour is related with IQ.
- Self Control:
  - Roy Baumeister at http://www.psy.fsu.edu/~baumeistertice/

3. Neuroticism

Neuroticism is “characterized by a chronic level of emotional instability and proneness to psychological distress.” (APA Dictionary)

It describes the tendency to feel negative emotions, such as anxiety, fear, sadness, and hostility, particularly when under stress.

Facets empirically loading on neuroticism are self efficacy and locus of control. Although some literature considers these measures as independent traits of personality, they are much correlated with each other (with a correlation of about r=.6), as they both refer to the belief to have outcomes under control. Perceived Self Efficacy (Bandura, A.) is the “Subjective perception of own capability of performance or ability to attain results”. Self Efficacy is case dependent (one can have high self efficacy in one field but low in another). It is emphasized that self efficacy is different from ability. For example, some studies have documented that, given ability, girls have lower measure of self efficacy than boys (Pajares, 1996, among others). Locus of Control is the “perception of how much control individuals have over conditions of their lives”.

Locus of Control can be internal or external. If internal, the individual attributes events to his own control: success or insuccess is a consequence of his or her actions. If external, the individual believes instead that outcomes are a consequence of good or bad luck.

Measures for perceived self efficacy and locus of control:

Perceived Self efficacy:

Bandura does not endorse and trait specific scales, but generalized self efficacy scales can be found into the “Self Efficacy Measures” paragraph of the website on self efficacy:
http://www.des.emory.edu/mfp/self-efficacy.html

Locus of Control:
Rotter has a 29 items questionnaire to measure locus of control, available at http://wilderdom.com/psychology/loc/RotterLOC29.html

For reading on the facets of self efficacy and locus of control:
- All of Bandura’ publications can be found in his website: http://www.des.emory.edu/mfp/banpubs.html
- Website on self efficacy: http://www.des.emory.edu/mfp/self-efficacy.html

4. Agreeableness
“The tendency to act in a cooperative, unselfish manner, construed as one end of a dimension of individual differences (agreeableness vs. disagreeableness) in the Big Five personality model.” This dimension includes facets such as trust and compliance. (APA Dictionary)

5. Extraversion
“An orientation of one’s interests and energies toward the outer world of people and things rather than the inner world of subjective experience. Extraversion is a broad personality trait and, like introversion, exists on a continuum of attitudes and behaviors. Extroverts are relatively more outgoing, gregarious, sociable, and openly expressive.” (APA dictionary)
Roberts (2006) suggests that there are two aspects of Extraversion: Social Dominance and Social Vitality.
- Social Dominance: describes dominance, independent, and self-confidence, especially in social settings. See NEO PI for a measure.
- Social Vitality: describes sociability, positive effect, and gregariousness. See NEO PI and CPI (California Psychological Inventory) sociability scale for a measure.
A third facet is empirically related to both extraversion and conscientiousness:
- Sensation Seeking: “the tendency to search out and engage in thrilling activities as a method of increasing stimulation and arousal. It takes the form of engaging in highly stimulating activities accompanied by a perception of danger” (APA Dictionary). Zuckermann is the main author for sensation seeking: he is also the creator of a sensation seeking scale:
Zuckermann has recently published a book on sensation seeking and risky behaviour:

a2. POSITIVE AFFECT, WELL BEING AND HAPPINESS
“Positive effect is the internal feeling state that occurs when a goal has been attained, a source if threat has been avoided or the individual is satisfied with the present state of affairs”
Positive affect includes emotions like joy, contentment, and pride. Negative affect, which is only moderately inversely correlated with positive emotions, includes fear, anxiety, sadness, etc. Life satisfaction (Diener, Seligman) is a cognitive, not affective, appraisal of the quality of one’s life. Well-being is the presence of positive effect, the absence of negative and positive life satisfaction.

Measures:
- PANAS, “Positive and Negative Affect Schedule” (Watson, Clark, & Tellegen) for positive and negative affect. For presentation and for example of PANAS see: http://www.psychology.uiowa.edu/Faculty/Watson/PANAS-X.pdf

Literature:
- Websites, with links to publications:
  - http://www.authentichappiness.sas.upenn.edu/
  - http://www.centreforconfidence.co.uk/pp/contributors.php
  - http://www.enpp.org/

a3. SELF ESTEEM

“The degree to which the qualities and characteristics contained in one’s self concept are perceived to be positive”
Self-esteem is a positive or negative orientation toward oneself: an overall evaluation of one's worth or value
It is a construct that enjoyed tremendous popularity in the 1970s but since has been considered epiphenomenal not causal. A minority of psychologists consider positive and negative evaluations of the self to be the sixth and seventh factors of personality.
It is measured with Self Esteem Rosenberg Scale. Link at: http://www.atkinson.yorku.ca/~psyctest/rosenbarg.pdf

Literature:

a4. TEMPERAMENT (childhood)

“Basic foundation of personality, usually assumed to be biologically determined and present early in life. It includes characteristics such as energy level, emotional responsiveness, response tempo and willingness to explore” (APA Dictionary)
There are fewer temperament variables than personality traits because these individual differences are less salient and stable in children than in adults.
After the pioneering work of Thomas and Chess (1977), different taxonomies have been proposed in the literature. A recent and very comprehensive is the one in Shiner R. And Caspi A. (2003), which individuates four main domains in child personality:
1) Extraversion/Positive Emotionality, with lower order traits of social inhibition, sociability, dominance and activity level. 2) Neuroticism/Negative Emotionality with lower traits of anxious distress (tapping fear and anxiety) and irritable distress (tapping anger and irritability).
3) Conscientiousness/Constraints, with lower traits of attention, inhibitory control and achievement motivation. 4) Agreeableness
Little is still known about how these early emerging differences evolve into personality traits.

Measures:
There are different typologies of measures for children personality. It should be emphasized that all of them have biases, and that therefore is always good to use more than one. The typologies are, as indicated in Shiner and Caspi (2003):

1) **Naturalistic observations** (infants): Naturalistic observations are based on the direct observation of the child’s behavior in naturalistic settings, such as the child's home environment. In this procedure, observers typically watch the child for periods of several hours and then use coding procedures for observing a variety of specific behavioral tendencies. The most used behavioural code is Eaton, Enns and Presse (1987), Journal of Psychoeducational Assessment 3, 273-280. The code is available on the article. Some rate the child’s temperament across a variety of dimensions after the observation is completed, using one of the standard temperament questionnaires.

2) **Questionnaires** completed by parents or teachers (infants and children):
These questionnaires mainly study temperament along the lines of the “big five”
- IBQ (Infant Behavior Questionnaire): M. Rothbart
Principal component analyses at the item level indicated that, for each age level, the first five principal components tended to group items according to the Big Five: five domains: Conscientiousness, Benevolence, Extraversion, Imagination, and Emotional Stability.
- Goldberg (2001):”Analyses of Digman’s child-personality data: Derivation of Big Five factor scores from each of six samples. Journal of Personality, 69, 709–744. Working on Digman data on Hawaiian children shows that the big five are derived as dimensions of personality of children.

3) **Q-Sets** (children): an informant (parents, teachers, observers or clinicians) sorts a set of cards into a quasi normal distribution based on how well each item describes the child. Also constructed along the “big five” taxonomy.
The main two Q-Sets are:
- California Child Q-Set: Block J., Block JH.(1969/1980): “The California Child Q-Set” Palo Alto, CA: Consulting Psychologists Press. [Not publicly available](http://www.bowdoin.edu/~sputnam/rothbart-temperament-questionnaires/). This Q-set consists of 100 cards; on each was a descriptive statement which parents rank-ordered into nine categories ranging from "most descriptive" to "least descriptive" of their child.

4) **Laboratory tasks** (children):
- Laboratory procedures (LAB-TAB: Goldsmith, Rothbart, 1993): contains measures such as Stranger Approach, Modified Peek-a-Boo game, and Puppet Game. In the lab, 20 episodes or games are used to elicit reactions of frustration (anger), wariness (fear), interest, pleasure, and activity level.
- Preschool Lab-TAB (Goldsmith, Reilly, 1995):
  Information on both at:
  http://psych.wisc.edu/goldsmith/Researchers/GEO/lab_TAB.htm
  The LAB-TAB manual can be downloaded at:
  http://psych.wisc.edu/goldsmith/Researchers/GEO/Lab_TAB_download_info.htm
  Some of the studies using LAB-TAB:

5) **Peer Nominations** (adolescents): the peer group nominates who is the best described by a particular item.

6) **Self Reports** (adolescents, only recently also children).

**Literature:**
- Shiner, R. (Personality Differences in Childhood and adolescence: measurement, development and consequences”, Journal of Child Psychology and Psychiatry, 44 (1), 2003, p.2-32
- Jerry Kagan:
- Mary Rothbart:
- Rothbart, M, Bates, J (1998)”Temperament”, in Handbook of child psychology: Vol 3, Social, emotional and personality development, Damon Eisenberg Eds,

**a5. PSYCHOPATHOLOGY**

A broad category comprising dysfunctional patterns of thought, feeling, or behavior. Most disorders are included in the DSM-IV manual. Axis I disorders (e.g., depression) are more intense and episodic/discreet, whereas Axis II disorders (i.e., personality disorders) are more tonic and enduring.

**Measure:**
For depression, the main scale is the Beck one. It indicates the acuteness of depression, but it is only for adults.
- Beck, A., (1961) ”An inventory for measuring depression “, Arch Gen Psychiatry, Archives of General Psychiatry 4, 561-571
A version for kids is Achenbach Child Behavior Checklist:
b) Abilities

b1. IQ AND “G” FACTOR

“Ability to reason, solve problems, comprehend abstract associations, learn from experience and think abstractly.”

One dominant factor “g” (“general factor”) refers to the first factor extracted from a factor analysis of cognitive tasks, which many researchers consider to represent general (vs. specific) intelligence. “It represents individuals’ abilities to perceive relationships and to derive conclusions from them. It is the basic ability that underlies the performance of different intellectual tasks”

IQ specifically refers to one’s intelligence relative to one’s age group (especially for children)

Main authors: Arthur Jensen; Nathan Brody

Measures:
- WISC: Wechsler Intelligence Scale for Children
- WAIS
- Raven’s Progressive Matrices
- ASVAB

Literature:

b2. SPECIFIC MENTAL ABILITIES

“Abilities as measured by tests of an individual in areas of spatial visualization, perceptual need, number facility, verbal comprehension, word fluency, memory, inductive reasoning and so forth”

An umbrella category for lower-level mental abilities, including math, verbal, and spatial abilities, as well as even more specific mental capacities. Measured by subtest scores on IQ tests


b3. CREATIVITY

Ability to generate novel ideas and behaviors that solve problems.

Measures:
- One measure for creativity is Gough 1979 “Creative Personality Scale”:

Literature:
b4. EXECUTIVE FUNCTION

“Higher level cognitive processes that organize and order behavior, including logic and reasoning, abstract thinking, problem solving, planning and carrying out and terminating goal directed behavior.”

Broad set of higher-level cognitive capacities attributed to the prefrontal cortex, often involving the coordination and management of lower-level processes” (APA Dictionary)

Executive functions are unrelated to “g”, but determine more basic abilities that can vary considerably among the population and that can be important in the process of acquiring cognitive skills: for example, Kim, Whyte, Vaccaro et al. show how executive functions may relate to attention, while Carpenter shows the relationship with working memory.

Executive functions are usually measured by neuropsychology tasks (e.g., go/no-go, Stroop, Continuous Performance Task)

Literature:
- Miller Lab, with links to publications: http://www.millerlab.org

b6. EMOTIONAL INTELLIGENCE

“Ability to process emotional information and use it in reasoning and other cognitive activities. According to Mayer and Slovey 1997 model It comprises four abilities: to perceive and appraise emotions accurately, to access and evoke emotions when they facilitate cognition, to comprehend emotional language and make use of emotional information, and to regulate one’s own and others’ emotions to promote growth and well-being” Ability to perceive emotion, to integrate it in thought, to understand it and to manage it (Mayer)” (APA Dictionary)

Emotion is here considered as a feeling state that conveys information about relationships (Mayer), and intelligence refers to the ability of reason validly about this information.

Measures:
The most reliable measure is considered to be the MSCEIT test, but it is not publicly available.
The validity of self reported judgments is still highly debated

Literature:
- John Mayer website for emotional intelligence, with links to all relevant papers: http://www.unh.edu/emotional_intelligence/

Goleman D.
### Motivation and IQ Summary

Many of these studies used a within-subject design and measured the effect of motivation as the differential between performance with and without incentives. An intriguing possibility for measuring motivation not available at the time of these early studies is offered by Pailing & Segalowitz (2004). A particular error-related ERP (event-related potential) is a metric for the salience or importance of making an error. Subjects who are high in Conscientiousness or low in Neuroticism showed dampened changes in this particular ERP when incentives were given for improved performance.

Table 3. Studies documenting an effect of motivation at the time of test administration and IQ score

<table>
<thead>
<tr>
<th>Study</th>
<th>Sample and Study Design</th>
<th>Experimental Group</th>
<th>Effect size of incentive (in standard deviations)</th>
<th>Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Edlund (1972)</td>
<td>Between subjects study. 11 matched pairs of low SES children; children were about 1 SD below average in IQ at baseline</td>
<td>M&amp;M candies given for each right answer</td>
<td>Experimental group scored 12 points higher than control group during a second testing on an alternative form of the Stanford Binet (about .8 SD)</td>
<td>“…a carefully chosen consequence, candy, given contingent on each occurrence of correct responses to an IQ test, can result in a significantly higher IQ score.”(p. 319)</td>
</tr>
<tr>
<td>Ayllon &amp; Kelly (1972) Sample 1</td>
<td>Within subjects study. 12 mentally retarded children (avg IQ 46.8)</td>
<td>Tokens given in experimental condition for right answers exchangeable for prizes</td>
<td>6.25 points out of a possible 51 points on Metropolitan Readiness Test. $t = 4.03$</td>
<td>“…test scores often reflect poor academic skills, but they may also reflect lack of motivation to do well in the criterion test…These results, obtained from both a population typically limited in skills and ability as well as from a group of normal children (Experiment II), demonstrate that the use of reinforcement procedures applied to a behavior that is tacitly...</td>
</tr>
<tr>
<td>Ayllon &amp; Kelly (1972) Sample 2</td>
<td>Within subjects study 34 urban fourth graders (avg IQ = 92.8)</td>
<td>Tokens given in experimental condition for right answers exchangeable for prizes</td>
<td>$t = 5.9$</td>
<td></td>
</tr>
<tr>
<td>Ayllon &amp; Kelly (1972) Sample 3</td>
<td>Within subjects study of 12 matched pairs of mentally retarded children</td>
<td>Six weeks of token reinforcement for good academic performance</td>
<td>Experimental group scored 3.67 points out of possible 51 points on a post-test given</td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Design</td>
<td>Subjects</td>
<td>Treatment</td>
<td>Results</td>
</tr>
<tr>
<td>-------</td>
<td>--------</td>
<td>----------</td>
<td>-----------</td>
<td>---------</td>
</tr>
<tr>
<td>Clingman &amp; Fowler (1976)</td>
<td>Within subjects study of 72 first- and second-graders assigned randomly to contingent reward, noncontingent reward, or no reward conditions.</td>
<td>M&amp;Ms given for right answers in contingent cdtn; M&amp;Ms given regardless of correctness in noncontingent cdtn</td>
<td>Only among low-IQ (&lt;100) subjects was there an effect of the incentive. Contingent reward group scored about .33 SD higher on the Peabody Picture Vocab test than did no reward group.</td>
<td>“…contingent candy increased the I.Q. scores of only the ‘low I.Q.’ children. This result suggests that the high and medium I.Q. groups were already functioning at a higher motivational level than children in the low I.Q. group.”</td>
</tr>
<tr>
<td>Zigler &amp; Butterfield (1968)</td>
<td>Within and between subjects study of 40 low SES children who did or did not attend nursery school were tested at the beginning and end of the year on Stanford-Binet Intelligence Test under either optimized or standard cdtns.</td>
<td>Motivation was optimized without giving test-relevant information. Gentle encouragement, easier items after items were missed, etc.</td>
<td>At baseline (in the fall), there was a full standard deviation difference (10.6 points and SD was about 9.5 in this sample) between scores of children in the optimized vs standard cdtns. The nursery group improved their scores, but only in the standard condition.</td>
<td>“…performance on an intelligence test is best conceptualized as reflecting three distinct factors: (a) formal cognitive processes; (b) informational achievements which reflect the content rather than the formal properties of cognition, and (c) motivational factors which involve a wide range of personality variables. (p. 2) “…the significant difference in improvement in</td>
</tr>
<tr>
<td>Source</td>
<td>Study Design</td>
<td>Participants</td>
<td>Interventions</td>
<td>Outcomes</td>
</tr>
<tr>
<td>--------</td>
<td>--------------</td>
<td>--------------</td>
<td>---------------</td>
<td>----------</td>
</tr>
<tr>
<td>Breuning &amp; Zella (1978)</td>
<td>Within and between subjects study of 485 special education high school students all took IQ tests, then were randomly assigned to control or incentive groups to retake tests. Subjects were below-average in IQ.</td>
<td>Incentives such as record albums, radios (&lt;$25) given for improvement in test performance</td>
<td>Scores increased by about 17 points. Results were consistent across the Otis-Lennon, WISC-R, and Lorge-Thorndike tests.</td>
<td>“In summary, the promise of individualized incentives on an increase in IQ test performance (as compared with pretest performance) resulted in an approximate 17-point increase in IQ test scores. These increases were equally spread across subtests. The incentive condition effects were much less pronounced for students having pretest IQs between 98 and 120 and did not occur for students having pretest IQs between 121 and 140.” (p. 225)</td>
</tr>
<tr>
<td>Holt &amp; Robbs (1979)</td>
<td>Between and within subjects study of 80 delinquent boys randomly assigned to 3 experimental groups and 1 control group. Each exp group received a standard and modified administration of the WISC-verbal section.</td>
<td>Exp 1-Token reinforcement for correct responses; Exp 2 - Tokens forfeited for incorrect responses (punishment), Exp 3-feedback on correct/incorrect responses</td>
<td>1.06 standard deviation difference between the token reinforcement and control groups (inferred from t = 3.31 for 39 degrees of freedom)</td>
<td>“Knowledge of results does not appear to be a sufficient incentive to significantly improve test performance among below-average I.Q. subjects...Immediate rewards or response cost may be more effective with below-average I.Q. subjects while other conditions may be more effective with average or above-average subjects.” (p. 83)</td>
</tr>
<tr>
<td>Larson (1994)</td>
<td>Between subjects study of 109 San Diego State University psychology students</td>
<td>Up to $20 for improvement over baseline performance on cognitive speed tests</td>
<td>“While both groups improved with practice, the incentive group improved slightly more.” → need to calculate effect size, but it was not large</td>
<td>2 reasons why incentive did not produce dramatic increase: 1) few or no unmotivated subjects among college volunteers, 2) information processing tasks are too simple for ‘trying harder’ to matter</td>
</tr>
<tr>
<td>Duckworth (in preparation)</td>
<td>Within subjects study of 61 urban low-achieving high school students tested with a group-administered Otis-Lennon IQ test during their freshman year, then again 2 years later with a one-on-one (WASI) test</td>
<td>Standard directions for encouraging effort were followed for the WASI brief test. Performance was expected to be higher because of the one-on-one environment.</td>
<td>Performance on the WASI as juniors was about 16 points higher than on the group-administered test as freshmen. Notably, on the WASI, this population looks almost “average” in IQ, whereas by Otis-Lennon standards they are low IQ. ( t(60) = 10.67, p &lt; .001 )</td>
<td>The increase in IQ scores could be attributed to any combination of the following 1) an increase in “g” due to schooling at an intensive charter school, 2) an increase in knowledge or crystallized intelligence, 3) an increase in motivation due to the change in IQ test format, and/or 4) an increase in motivation due to experience at high performing school</td>
</tr>
</tbody>
</table>


Ability Bias, Errors in Variables and Sibling Methods

James J. Heckman
University of Chicago

Econ 312
This draft, May 26, 2006
1 Ability Bias

Consider the model:

$$\log y_{it} = \alpha_0 + \alpha_1 S_i + U_{it}$$

where $y_{it} =$ income, $S_i =$ schooling, and $\alpha_0$ and $\alpha_1$ are parameters of interest. What we have omitted from the above specification is unobserved ability, which is captured in the residual term $U_{it}$. We thus re-write the above as:

$$\log y_{it} = \alpha_0 + \alpha_1 S_i + a_i + \varepsilon_{it}$$

where $a_i$ is ability, $(\varepsilon_{it}, \varepsilon_{i't}) \perp (S_i, S_{i'})$, and we believe that $Cov(a_i, S_i) \neq 0$. Thus, $E(U_{it} \mid S_i) \neq 0$, so that OLS on our original specification gives biased and inconsistent estimates.
1.1 Strategies for Estimation

1. *Use proxies for ability*: Find proxies for ability and include them as regressors. Examples may include: height, weight, etc. The problem with this approach is that proxies may measure ability with error and thus introduce additional bias (see Section 1.3).
2. **Fixed Effect Method**: Find a paired comparison. Examples may include a genetic twin or sibling with similar or identical ability. Consider two individuals $i$ and $i'$:

$$
\log y_{it} - \log y_{i't} = (\alpha_0 + \alpha_1 S_i + U_{it}) - (\alpha_0 + \alpha_1 S_{i'} + U_{i't}) = \alpha_1 (S_i - S_{i'}) + (a_i - a_{i'}) + (\varepsilon_{it} - \varepsilon_{i't})
$$

Note: if $a_i = a_{i'}$, then OLS performed on our fixed effect
estimator is unbiased and consistent. If \( a_i \neq a_i' \), then we just get a different bias (see Section 1.2). Further, if \( S_i \) is measured with error, we may exacerbate the bias in our fixed effect estimator (see Section 1.3).

### 1.2 OLS vs. Fixed Effect (FE)

In the \textit{OLS} case with ability bias, we have:

\[
\text{plim } (\alpha_1^{OLS}) = \alpha_1 + \frac{\text{Cov}(a, S)}{\text{Var}(S)}
\]

(See derivation of Equation (2.2) for more background on the above derivation).
We also impose:

\[
\begin{align*}
Var(S) &= Var(S') \\
Cov(a, S) &= Cov(a', S') \\
Cov(a', S) &= Cov(a, S')
\end{align*}
\]

With these assumptions, our fixed effect estimator is given by:

\[
\plim \alpha_1^{FE} = \alpha_1 + \frac{Cov(S - S', (a - a') + (\varepsilon - \varepsilon'))}{Var(S - S')} = \alpha_1 + \frac{Cov(a, S) - Cov(a', S)}{Var(S) - Cov(S, S')}
\]

Note that if \(Cov(a', S) = 0\), and ability is positively correlated with schooling, then the fixed effect estimator is upward biased.
From the preceding, we see that the fixed effect estimator has more asymptotic bias if:

\[
\frac{\text{Cov}(a, S) - \text{Cov}(a', S)}{\text{Var}(S) - \text{Cov}(S, S'')} > \frac{\text{Cov}(a, S)}{\text{Var}(S)}
\]

\[
\Rightarrow \quad \frac{\text{Cov}(a, S)}{\text{Var}(S)} > \frac{\text{Cov}(a', S)}{\text{Cov}(S, S'')}
\]
1.3 Measurement Error

Say $S^* = S + \nu$, where $S^*$ is observed schooling. Our model now becomes:

$$\log y = \alpha_0 + \alpha_1 S + U = \alpha_0 + \alpha_1 S^* + (a + \varepsilon - \alpha_1 \nu)$$

and the fixed effect estimator gives:

$$\log y - \log y' = (\alpha_0 + \alpha_1 S + U) - (\alpha_0 + \alpha_1 S' + U')$$

$$= \alpha_1 (S^* - S') + (U - U') + \alpha_1 (\nu' - \nu)$$

Now we wish to examine which estimator (OLS or fixed effect), has more asymptotic bias given our measurement error problem. For the remaining arguments of this section, we assume:

$$E (\nu | S) = E (\nu' | S) = E (\nu | \nu') = 0$$

so that the OLS estimator gives:
\[ \text{plim } \alpha_1^{OLS} = \alpha_1 + \frac{\text{Cov}(S^*, a + \varepsilon - \alpha_1 \nu)}{\text{Var}(S^*)} \]
\[ = \alpha_1 + \frac{\text{Cov}(a, S) - \alpha_1 \text{Var}(\nu)}{\text{Var}(S) + \text{Var}(\nu)}. \]

The fixed effect estimator gives:

\[ \text{plim } \alpha_1^{FE} = \alpha_1 + \frac{\text{Cov} \left( S^* - S^* ', (U - U') + \alpha_1 (\nu' - \nu) \right)}{\text{Var}(S^* - S^*)} \]
\[ = \alpha_1 + \frac{\text{Cov} ((S - S'), (a - a')) - \alpha_1 \text{Var}(\nu' - \nu)}{\text{Var}(S - S') + \text{Var}(\nu' - \nu)} \]
\[ = \alpha_1 + \frac{\text{Cov}(a, S) - \text{Cov}(a, S') - \alpha_1 \text{Var}(\nu)}{\text{Var}(S) + \text{Var}(\nu) - \text{Cov}(S', S)} . \]
Under what conditions will the fixed effect bias be greater? From the above, we know that this will be true if and only if:

\[
\frac{\text{Cov}(a, S) - \text{Cov}(a, S') - \alpha_1 \text{Var}(\nu)}{\text{Var}(S) + \text{Var}(\nu) - \text{Cov}(S', S)} > \frac{\text{Cov}(a, S) - \alpha_1 \text{Var}(\nu)}{\text{Var}(S) + \text{Var}(\nu)}
\]

\[
\Rightarrow \text{Cov}(a, S') (\text{Var}(S) + \text{Var}(\nu)) > (\alpha_1 \text{Var}(\nu) - \text{Cov}(a, S)) \text{Cov}(S', S)
\]

\[
\Rightarrow \frac{\text{Cov}(a, S) - \alpha_1 \text{Var}(\nu)}{\text{Var}(S) + \text{Var}(\nu)} > \frac{\text{Cov}(a, S')}{\text{Cov}(S', S)}.
\]

If this inequality holds, taking differences can actually worsen the fit over OLS alone. Intuitively, we see that we have differenced out the true component, \(S\), and compounded our measurement error problem with the fixed effect estimator.
In the special case $a = a'$, the condition is

\[
\frac{-\alpha_1 \text{Var}(\nu)}{\text{Var}(S) + \text{Var}(\nu) - \text{Cov}(S', S)} > \frac{\text{Cov}(a, S) - \alpha_1 \text{Var}(\nu)}{\text{Var}(S) + \text{Var}(\nu)}.
\]
2 Errors in Variables

2.1 The Model

Suppose that the equation for earnings is given by:

\[ Y_t = X_{1t} \beta_1 + X_{2t} \beta_2 + U_t \]

where \( E(U_t \mid X_{1t}, X_{2t}) = 0 \ \forall \ t, t' \). Also define:

\[ X_{1t}^* = X_{1t} + \varepsilon_{1t} \quad \text{and} \quad X_{2t}^* = X_{2t} + \varepsilon_{2t}. \]
Here, $X_{1t}$ and $X_{2t}$ are observed and measure $X_{1t}$ and $X_{2t}$ with error. We also impose that $X_i \perp \varepsilon_j \forall i, j$. So, our initial model can be equivalently re-written as:

$$Y_t = X_{1t}^* \beta_1 + X_{2t}^* \beta_2 + (U_t - \varepsilon_1 \beta_1 - \varepsilon_2 \beta_2).$$

Finally, by assumed independence of $X$ and $\varepsilon$, we write:

$$\Sigma_{x^*} = \Sigma_x + \Sigma_\varepsilon.$$
2.2 McCallum’s Problem

Question: Is it better for estimation of $\beta_1$ to include other variables measured with error? Suppose that $X_{1t}$ is not measured with error, in the sense that $\varepsilon_{1t} = 0$, while $X_{2t}$ is measured with error. In 2.2.1 and 2.2.2 below, we consider both excluding and including $X_{2t}$, and investigate the asymptotic properties of both cases.

2.2.1 Excluded $X_{2t}$

The equation for earnings with omitted $X_2$ is:

$$y = X_1 \beta_1 + (U + X_2 \beta_2)$$
Therefore, by arguments similar to those in the appendix, we know:

\[
\text{plim } \tilde{\beta}_1 = \beta_1 + \frac{\sigma_{12}}{\sigma_{11}} \beta_2. \tag{2.1}
\]

Here, \(\sigma_{12}\) is the covariance between the regressors, and \(\sigma_{11}\) is the variance of \(X_1\). Before moving on to a more general model for the inclusion of \(X_{2t}\), let us first consider the classical case for including both variables. Suppose

\[
\Sigma_\varepsilon = \begin{bmatrix} \sigma_{11}^* & 0 \\ 0 & \sigma_{22}^* \end{bmatrix}, \Sigma_x = \begin{bmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{22} \end{bmatrix}.
\]

We know that:

\[
\text{plim } \hat{\beta} = \left[ I - (\Sigma_x^*)^{-1} (\Sigma_\varepsilon) \right] \beta \tag{2.2}
\]
where the coefficient and regressor vectors have been stacked appropriately (see Appendix for derivation). Note that $\Sigma_e$ represents the variance-covariance matrix of the measurement errors, and $\Sigma_x$ is the variance-covariance matrix of the regressors. Straightforward computations thus give:

$$\text{plim } \hat{\beta}$$

$$= \begin{bmatrix}
I - \begin{bmatrix}
\sigma_{11} + \sigma_{11}^* \\
0 \\
\sigma_{22} + \sigma_{22}^*
\end{bmatrix}^{-1} \begin{bmatrix}
\sigma_{11}^* \\
0 \\
\sigma_{22}^*
\end{bmatrix} & \begin{bmatrix}
\beta_1 \\
\beta_2
\end{bmatrix} \\
\frac{\sigma_{11}}{\sigma_{11} + \sigma_{11}^*} & 0 \\
0 & \frac{\sigma_{22}}{\sigma_{22} + \sigma_{22}^*}
\end{bmatrix}$$

$$= \begin{bmatrix}
\beta_1 \\
\beta_2
\end{bmatrix}.$$
2.2.2 Included $X_{2t}$

In McCallum’s problem we suppose that $\sigma_{12}^* = 0$. Further, as $X_{1t}$ is not measured with error, $\sigma_{11}^* = 0$. Substituting this into equation 2.2 yields:

$$\text{plim} \hat{\beta} = \beta - \left[ \begin{array}{cc} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} + \sigma_{22}^* \end{array} \right]^{-1} \left[ \begin{array}{c} 0 \\ 0 \end{array} \right] \beta$$

With a little algebra, the above gives:

$$\text{plim} \hat{\beta}_1 = \beta_1 + \beta_2 \left( \frac{\sigma_{12}}{\sigma_{11}} \right) \left( \frac{\sigma_{22}^*}{\sigma_{22} + \sigma_{22}^* - \frac{\sigma_{12}^2}{\sigma_{11}}} \right) \left( \frac{\sigma_{22}}{\sigma_{22} (1 - \rho_{12}^2) + \sigma_{22}^*} \right)$$
where $\rho_{12}^2$ is simply the correlation coefficient, $\frac{\sigma_{12}^2}{\sigma_{11}\sigma_{22}}$. Further, we know that:

$$0 < \rho_{12}^2 < 1$$

so including $X_{2t}$ results in less asymptotic bias (inconsistency). (We get this result by comparing the above with the bias from excluding $X_{2t}$ in section 2.2.1, the result captured in equation (2.1)). So, we have justified the kitchen sink approach. This result generalizes to the multiple regressor case - 1 badly measured variable with $k$ good ones (Econometrica, 1972).
2.3 General Case

In the most general case, we have:

\[
\text{plim } \hat{\beta} = \beta - (\Sigma_{x^*})^{-1} \sum_{\varepsilon} \beta
\]

\[
= \beta - \left[ \begin{array}{cc}
\sigma_{11} + \sigma_{11}^* & \sigma_{12} + \sigma_{12}^* \\
\sigma_{12} + \sigma_{12}^* & \sigma_{22} + \sigma_{22}^*
\end{array} \right]^{-1} \left[ \begin{array}{cc}
\sigma_{11}^* & \sigma_{12}^* \\
\sigma_{12}^* & \sigma_{22}^*
\end{array} \right] \left[ \begin{array}{c}
\beta_1 \\
\beta_2
\end{array} \right].
\]

With a little algebra we find:

\[
\det(\Sigma_{x^*}) = \sigma_{11}\sigma_{22} + \sigma_{11}\sigma_{22}^* + \sigma_{11}\sigma_{22} + \sigma_{11}\sigma_{22}^* - \sigma_{12}^2 - 2\sigma_{12}\sigma_{12}^* - \sigma_{12}^2
\]
Therefore:

$$\text{plim } \hat{\beta} = \beta - \frac{1}{\det(\Sigma_{x^*})} \begin{bmatrix} \sigma_{22} + \sigma_{22}^* & - (\sigma_{12} + \sigma_{12}^*) \\ - (\sigma_{12} + \sigma_{12}^*) & \sigma_{11} + \sigma_{11}^* \end{bmatrix}$$

$$\times \begin{bmatrix} \sigma_{11}^* & \sigma_{12}^* \\ \sigma_{12}^* & \sigma_{22}^* \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$$

Supposing $$\sigma_{12}^* = 0$$, we get:

$$
\det(\tilde{\Sigma}_{x^*}) = \det(\Sigma_{x^*}) \mid_{\sigma_{12}^* = 0} \\
= \sigma_{11}\sigma_{22} + \sigma_{11}\sigma_{22}^* + \sigma_{11}\sigma_{22} + \sigma_{11}\sigma_{22}^* - \sigma_{12}^2
$$
and thus:

\[
\text{plim } \hat{\beta} = \beta - \begin{bmatrix}
\frac{(\sigma_{22} + \sigma_{22}^*)\sigma_{11}^*}{\det(\Sigma_{x^*})} & \frac{-\sigma_{12}\sigma_{22}^*}{\det(\Sigma_{x^*})} \\
\frac{-\sigma_{11}^*\sigma_{12}}{\det(\Sigma_{x^*})} & \frac{(\sigma_{11} + \sigma_{11}^*)\sigma_{22}^*}{\det(\Sigma_{x^*})}
\end{bmatrix} \begin{bmatrix}
\beta_1 \\
\beta_2
\end{bmatrix}
\]

Note that if \( \beta_2\sigma_{12} < 0 \), \( OLS \) may not be downward biased for \( \beta_1 \). If \( \beta_2 = 0 \), we get:

\[
\text{plim } \hat{\beta}_2 = \frac{\beta_1\sigma_{12}\sigma_{11}^*}{\det(\tilde{\Sigma}_{x^*})}
\]

so, if \( X_2 \) were a race variable and blacks get lower quality schooling, (where schooling is measured by \( X_{1t} \),) then \( \sigma_{12} < 0 \), and hence \( \hat{\beta}_2 < 0 \). This would be a finding in support of labor market discrimination.
2.4 The Kitchen Sink Revisited

McCallum’s analysis suggests that one should toss in a variable measured with error if there is no measurement error in $X_{1t}$. But suppose that there is measurement error in $X_{1t}$. Is it still better to include the additional variable measured with error as a regressor? We proceed by imposing $\beta_2 = 0$.

(i) **Excluded** $X_{2t}$. The equation for earnings with measurement error in $X_1$ and excluded $X_2$ is:

$$y = (X_1^* + \varepsilon_1) \beta_1 + (U + X_2 \beta_2)$$

$$= X_1^* \beta_1 + (U + X_2 \beta_2 + \beta_1 \varepsilon_1)$$
Therefore:

\[
\text{plim } \hat{\beta}_1 = \beta_1 - \beta_1 \left( \frac{\sigma_{11}^*}{\sigma_{11} + \sigma_{11}^*} \right) = \beta_1 \left( \frac{\sigma_{11}}{\sigma_{11} + \sigma_{11}^*} \right) = \beta_1 \left( \frac{1}{1 + \frac{\sigma_{11}^*}{\sigma_{11}}} \right)
\]  

(2.3)

(ii) Included \( X_{2t} \). From our analysis in the General Case (Section 2.3), we know that:

\[
\text{plim } \hat{\beta}_1 = \beta_1 \left( \frac{(\sigma_{22} + \sigma_{22}^*) \sigma_{11} - \sigma_{12}^2}{\det(\tilde{\Sigma}_{x^*})} \right).
\]  

(2.4)
If $\sigma_{22}^* = 0$, so that $X_{2t}$ is not measured with error:

$$\text{plim } \hat{\beta}_1 = \beta_1 \left( \frac{\sigma_{11} \sigma_{22} - \sigma_{12}^2}{\sigma_{11} \sigma_{22} - \sigma_{12}^2 + \sigma_{11}^* \sigma_{22}} \right)$$

(2.5)

$$= \beta_1 \left( \frac{1 - \rho_{12}^2}{1 - \rho_{12}^2 + \frac{\sigma_{11}^*}{\sigma_{11}}} \right).$$

Comparing eqn (2.4) and eqn (2.5), we see that adding the variable measured without error always exacerbates the bias.
For, the bias in the excluded case will be smaller if:

\[
\beta_1 \left( \frac{1}{1 + \frac{\sigma_{11}^*}{\sigma_{11}}} \right) > \beta_1 \left( \frac{1 - \rho_{12}^2}{1 - \rho_{12}^2 + \frac{\sigma_{11}^*}{\sigma_{11}}} \right)
\]

\[\leftrightarrow \left( 1 - \rho_{12}^2 + \frac{\sigma_{11}^*}{\sigma_{11}} \right) > \left( 1 + \frac{\sigma_{11}^*}{\sigma_{11}} \right) (1 - \rho_{12}^2)\]

\[\leftrightarrow 0 > -\rho_{12}^2 \frac{\sigma_{11}^*}{\sigma_{11}}.\]

which is always the case, provided \( \rho_{12}^2 > 0 \). (Note that the coefficients on \( \beta_1 \) for both the excluded and included case are less than one. So, the larger coefficient is the one with less bias, as stated above.)
Now suppose that $\sigma_{22}^* > 0$, so that both variables are measured with error. Then:

$$\text{plim } \hat{\beta}_1 = \beta_1 \left( \frac{(\sigma_{22} + \sigma_{22}^*) \sigma_{11} - \sigma_{12}^2}{\det(\tilde{\Sigma}_{x^*})} \right)$$

$$= \beta_1 \left( 1 + \frac{\sigma_{22}^*}{\sigma_{22}} - \rho_{12}^2 \right)$$

Intuitively, adding measurement error in $X_{2t}$ can only worsen the bias, and thus exclusion should again be preferred to inclusion. Formally, including $X_{2t}$ gives more bias if and only
if:

\[
\beta_1 \left( \frac{1 + \frac{\sigma_{22}^*}{\sigma_{22}} - \rho_{12}^2}{1 + \frac{\sigma_{11}^*}{\sigma_{11}} + \frac{\sigma_{11}^* \sigma_{22}^*}{\sigma_{11} \sigma_{22}} + \frac{\sigma_{22}^*}{\sigma_{22}} - \rho_{12}^2} \right) < \beta_1 \left( \frac{1}{1 + \frac{\sigma_{11}^*}{\sigma_{11}}} \right)
\]

\[\iff \left( 1 + \frac{\sigma_{11}^*}{\sigma_{11}} \right) \left( 1 + \frac{\sigma_{22}^*}{\sigma_{22}} - \rho_{12}^2 \right) < \left( 1 + \frac{\sigma_{11}^*}{\sigma_{11}} + \frac{\sigma_{11}^* \sigma_{22}^*}{\sigma_{11} \sigma_{22}} + \frac{\sigma_{22}^*}{\sigma_{22}} - \rho_{12}^2 \right)\]

\[\iff -\rho_{12}^2 \frac{\sigma_{11}^*}{\sigma_{11}} < 0.\]
Thus, provided $\rho_{12}^2 > 0$, including $X_{2t}$ results in more bias than excluding it. If $\rho_{12}^2 = 0$, the bias from including $X_{2t}$ is obviously seen to be:

$$
\beta_1 \left( \frac{1 + \frac{\sigma_{22}^*}{\sigma_{22}^*}}{1 + \frac{\sigma_{11}^*}{\sigma_{11}} + \frac{\sigma_{11}^* \sigma_{22}^*}{\sigma_{11} \sigma_{22}^*} + \frac{\sigma_{22}^*}{\sigma_{22}^*}} \right) = \beta_1 \left( \frac{1 + \frac{\sigma_{22}^*}{\sigma_{22}^*}}{1 + \frac{\sigma_{22}^*}{\sigma_{22}^*}} \right) \left( 1 + \frac{\sigma_{11}^*}{\sigma_{11}} \right)
$$

$$
= \beta_1 \left( \frac{1}{1 + \frac{\sigma_{11}^*}{\sigma_{11}}} \right)
$$

so that including and excluding $X_{2t}$ yields the same result.
Finally, from the General Case section, we have:

\[
\operatorname{plim} \hat{\beta}_1 = \frac{\beta_1 (\sigma_{22} + \sigma_{22}^*) \sigma_{11} - \sigma_{12}^2 + \beta_2 (\sigma_{12} \sigma_{22}^*)}{\sigma_{11} \sigma_{22} - \sigma_{12}^2 + \sigma_{11}^* \sigma_{22}^* + \sigma_{11}^* \sigma_{22} + \sigma_{11} \sigma_{22}^*}.
\]

L’Hôpital’s rule on the above shows that:

\[
\begin{align*}
\sigma_{11}^* & \quad \longrightarrow \quad \infty \lim \left( \operatorname{plim} \hat{\beta}_1 \right) = 0, \quad \text{and} \\
\lim_{\sigma_{22}^* \to \infty} \left( \operatorname{plim} \hat{\beta}_1 \right) & = \frac{\beta_1 \sigma_{11} + \beta_2 \sigma_{12}}{\sigma_{11} + \sigma_{11}^*} \\
& = \frac{\beta_1 \sigma_{11}}{\sigma_{11} + \sigma_{11}^*} + \frac{\beta_2 \sigma_{12}}{\sigma_{11} + \sigma_{11}^*}.
\end{align*}
\]
Appendix

Derivation of Equation (2.2)
We can write

\[ y_t = x^* \beta + (U_t - \epsilon_{1t}\beta_1 - \epsilon_{2t}\beta_2), \]

where:

\[ x^* = \begin{bmatrix} x_1^* & x_2^* \end{bmatrix} \]

and \[ \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}, \]

and \( x_1^*, x_2^*, \) are \( T \times 1. \)
So:

\[ \hat{\beta}^{\text{OLS}} = (x^* x^*)^{-1} (x^* y) \]

\[ = \beta + \left( \frac{x^* x^*}{T} \right)^{-1} \left( x^* (U - \epsilon_1 \beta_1 - \epsilon_2 \beta_2) \right) \]

\[ = \beta + \left( \frac{x^* x^*}{T} \right)^{-1} \frac{x^*}{T} \left( U - \epsilon_1 \beta_1 - \epsilon_2 \beta_2 \right) \]

\[ \times \left( \left( \frac{x^* U}{T} \right) - \left( \frac{x^* \epsilon_1 \beta_1}{T} \right) - \left( \frac{x^* \epsilon_2 \beta_2}{T} \right) \right) \]

\[ \rightarrow \beta + \left[ E \left( x^* x^* \right) \right]^{-1} \]

\[ \times \left( E \left( x^* U \right) - E \left( x^* \epsilon_1 \right) \beta_1 - E \left( x^* \epsilon_2 \right) \beta_2 \right) \]
\[
\begin{align*}
\beta & = \beta - \left[ \begin{array}{cc}
E(x_1'^* x_1^*) & E(x_1'^* x_2^*) \\
E(x_2'^* x_1^*) & E(x_2'^* x_2^*)
\end{array} \right]^{-1} \\
& \quad \times \left( E \left[ \begin{array}{c}
x_1'^* \epsilon_1 \\
x_2'^* \epsilon_1
\end{array} \right] \beta_1 + E \left[ \begin{array}{c}
x_1'^* \epsilon_2 \\
x_2'^* \epsilon_2
\end{array} \right] \beta_2 \right) \\
\beta & = \beta - \left[ \begin{array}{cc}
E(x_1'^* x_1^*) & E(x_1'^* x_2^*) \\
E(x_2'^* x_1^*) & E(x_2'^* x_2^*)
\end{array} \right]^{-1} \\
& \quad \times \left[ \begin{array}{c}
E(x_1'^* \epsilon_1) \\
E(x_2'^* \epsilon_1)
\end{array} \right] \beta_1 \\
& \quad \times \left[ \begin{array}{c}
E(x_1'^* \epsilon_2) \\
E(x_2'^* \epsilon_2)
\end{array} \right] \beta_2
\end{align*}
\]
\[
= \left( I - (\Sigma_{x^*})^{-1} \right) \left[ \begin{array}{c}
E \left( (\varepsilon_1' + x_1') \epsilon_1 \right) \\
E \left( (\varepsilon_2' + x_2') \epsilon_1 \right)
\end{array} \right] \left[ \begin{array}{c}
E \left( (\varepsilon_1' + x_1') \epsilon_2 \right) \\
E \left( (\varepsilon_2' + x_2') \epsilon_2 \right)
\end{array} \right]
\times \begin{bmatrix}
\beta_1 \\
\beta_2
\end{bmatrix}
= \left( I - (\Sigma_{x^*})^{-1} (\Sigma_{\epsilon}) \right) \beta,
\]
where the second-to-last step follows from the independence of \( x \) and \( \varepsilon \). This type of argument is also used to derive the probability limit of the \( \beta \)'s in section 1.
3 Sibling Models: Components of Variance Scheme

Suppose that data on two brothers, say $\alpha$ and $\beta$, is at our disposal. Without loss of generality, we will consider how to estimate parameters of interest for person $\alpha$ in what follows. We will begin by introducing a general model and then focus on the two-person case mentioned above. Consider the following triangular system:

\[
\begin{align*}
y_{1ij} &= \varepsilon_{1ij} \\
y_{2ij} &= \nu_{12}y_{1ij} + \varepsilon_{2ij} \\
y_{3ij} &= \nu_{13}y_{1ij} + \nu_{23}y_{2ij} + \varepsilon_{3ij}
\end{align*}
\]
Here, $i, j$ indexes the $j^{th}$ person in the $i^{th}$ group. We assume that $\varepsilon_{lij}$ and $\varepsilon_{li'j}$ are uncorrelated (i.e., uncorrelated across groups). Further, we suppose:

$$\varepsilon_{kij} = \lambda_k h_{ij} + \mu_{kij}$$
$$h_{ij} = F_i + g_{ij},$$

for $k = 1, 2, 3$. We assume $\mu_{kij}$ is uncorrelated across equations and across $j$ within the group, $F_i$ is i.i.d. across groups, and $g_{ij}$ is i.i.d. within groups and uncorrelated with $F_i$. 
3.1 Estimation

We specialize the above model into a two person framework and propose a similar three equation system. Let $y_1 =$ early (pre-school) test score, $y_2 =$ schooling (years), and $y_3 =$ earnings. It seems plausible to write the equation system

$$
\begin{align*}
    y_1 & = h + U_1 \\
    y_2 & = \lambda_2 h + U_2. \\
    y_3 & = \nu_{23} y_2 + \lambda_3 h + U_3,
\end{align*}
$$

where $h =$ ability. Regressing $y_3$ on $y_2$ clearly gives biased estimates of $\nu_{23}$ as $E(h \mid y_2) \neq 0$. If $\lambda_3 > 0$, then OLS estimates of $\nu_{23}$ are upward biased. One estimation approach is to use $y_1$ as a proxy for ability:

$$
y_3 = \nu_{23} y_2 + \lambda_3 (y_1 - U_1) + U_3.
$$
However, this results in a similar problem — regressing $y_3$ on $y_1$ and $y_2$ will give biased estimates as $y_1$ is correlated with our residual. (i.e., $y_1$ is an imperfect proxy).

**Solutions:**

One solution is to use $y_{1\beta}$ as an instrument for $y_{1\alpha}$. Why is this a valid IV? From our construction of the model, we know that the $U_i$ are uncorrelated across equations and groups. Further, test scores are correlated across siblings. That is, $\text{Cov}(y_{1\alpha}, y_{1\beta}) \neq 0$ by our group structure.

Another solution is possible if there exists an additional early reading on the same person:

$$y_0 = \lambda_0 h + U_0.$$  

Then if $\lambda_0 \neq 0$, $y_0$ is a valid proxy for $y_1$, and we can perform 2SLS.
3.2 Griliches and Chamberlain model

Here we have a modified triangular system as follows:

\[\begin{align*}
y_1 &= \lambda_1 h + U_1 \\
y_2 &= \nu_{12} y_1 + \lambda_2 h + U_2 \\
y_3 &= \nu_{13} y_1 + \nu_{23} y_2 + \lambda_3 h + U_3
\end{align*}\]

where \(y_1\) = years schooling, \(y_2\) = late test score (SAT), and \(y_3\) = earnings. Note that there are alternative models with other dependent variables. For example, \(\{y_1 = \text{schooling}, y_2 = \text{early earnings}, \text{and } y_3 = \text{late earnings}\}\), and \(\{y_1 = \text{schooling}, y_2 = \text{consumption}, \text{and } y_3 = \text{earnings}\}\). Getting the equation system into reduced form and expressing as matrix notation, we write

\[y_k = d_k h + \rho_k,\]

\(1, 2, 3\)
where:

\[
d_k = \begin{bmatrix}
    d_1 \\
    d_2 \\
    d_3
\end{bmatrix} = \begin{bmatrix}
    \lambda_1 \\
    \lambda_2 + \nu_{12} \lambda_1 \\
    \lambda_3 + \nu_{13} \lambda_1 + \nu_{23}(\lambda_2 + \nu_{12} \lambda_1)
\end{bmatrix}
\]

and:

\[
\rho_k = \begin{bmatrix}
    \rho_1 \\
    \rho_2 \\
    \rho_3
\end{bmatrix} = \begin{bmatrix}
    \mu_1 \\
    \mu_2 + \nu_{12} \mu_1 \\
    \mu_3 + \nu_{13} \mu_1 + \nu_{23}(\mu_2 + \nu_{12} \mu_1)
\end{bmatrix}
\]

**Estimation.** For estimation, we impose that \(\nu_{23} = 0\). In our second example of section 3.2, this would be equivalent to stating that there is no correlation between transient income and consumption (permanent income hypothesis). In general, with one factor, we need one more exclusion than that implied by triangularity.
(i) $y_1$ proxies $h$.

$$h = \frac{y_1 - \rho_1}{d_1}$$

so that

$$y_2 = \frac{d_2}{d_1} y_1 - \frac{d_2}{d_1} \rho_1 + \rho_2.$$  

We can then estimate $\frac{d_2}{d_1}$ consistently by using $y_{1\beta}$ as an instrument for $y_{1\alpha}$ in the equation above.

(ii) Get residuals from (i): $z = \rho_2 - \frac{d_2}{d_1} \rho_1$.

(iii). Use the residuals as an instrument for $y_1$ in the $y_3$ equation. $Z$ is valid since it is both uncorrelated with $h$ and $U_3$, and it is correlated with $y_1$:  

40
\[
Cov(y_1, z) = Cov \left( y_1, \rho_2 - \frac{d_2}{d_1} \rho_1 \right)
\]
\[
= Cov \left( \lambda_1 h + U_1, U_2 + \nu_{12} U_1 - \frac{\lambda_2 + \nu_{12} \lambda_1}{\lambda_1} U_1 \right)
\]
\[
= Cov \left( \lambda_1 h + U_1, U_2 - \frac{\lambda_2}{\lambda_1} U_1 \right) \neq 0
\]

if \( U_1 \neq 0 \), and, \( \lambda_2 \neq 0 \). Thus we can estimate \( \nu_{13} \).

(iv). Interchange the role of \( y_2 \) and \( y_3 \) to estimate \( \nu_{12} \).

(v). Form the residual (and recall that \( \nu_{13} \) is known
and \( \nu_{23} = 0 \))

\[
w = y_3 - \nu_{13} y_1 = \lambda_3 h + U_3.
\]
(vi) Use $y_1$ as a proxy for ability. Substituting this into $V$ gives:

$$w = \frac{\lambda_3}{\lambda_1} y_1 + \frac{\lambda_3}{\lambda_1} U_3 - \frac{\lambda_3}{\lambda_1} U_1.$$ 

(vii) Now use $y_{1\beta}$ as an instrument for $y_{1\alpha}$ in the above to get an estimate of $\frac{\lambda_3}{\lambda_1}$.

(viii) Interchange the role of $y_2$ and $y_3$ to estimate $\frac{\lambda_2}{\lambda_1}$. 

42
3.3 Triangular systems more generally

Without loss of generality, suppose that $y_2$ is excluded from the $t^{th}$ equation of our system. (We are supposing the existence of an extra exclusion than that implied by triangularity). We seek to estimate the parameters of the system in equation $t$ as well as equations before and after $t$.

Equation $t$.

i. Use $y_1$ as a proxy for ability. Solving for $h$ and substituting into the equation:

$$y_k = d_k h + \rho_k$$
We get:

\[ y_k = \frac{d_k}{d_1} y_1 - \frac{d_k}{d_1} \rho_1 + \rho_k \]

and we are considering \( k = 2, \ldots, t - 1 \). The ratio \( \frac{d_k}{d_1} \) can then be identified using \( y_{1\beta} \) as an instrument for \( y_{1\alpha} \).

ii. Form the residuals:

\[ z_k = \rho_k - \frac{d_k}{d_1} \rho_1 \quad k = 2, \ldots t - 1 \]

Now we have \( t - 2 \) IV’s (\( z_2, z_3, \ldots, z_{t-1} \)) for the \( t - 2 \) independent variables in the \( t^{th} \) equation (\( y_1, y_3, \ldots, y_{t-1} \)), so we can consistently estimate the coefficients in the \( t^{th} \) equation.
Equations before t.

iii. Form:

\[ y_t^* = y_t - \nu_1 t y_1 - \cdots - \nu_{t-1} t y_{t-1} \]

We can use \( y_1, \cdots, y_{k-1}, y_t^* \) to form \( k - 1 \) purged IV’s and \( y_{t\beta}^* \) is used as a proxy for unobserved ability, \( h \). In this way, we can estimate all of the parameters in equations \( k < t \). (Note the sequential order implicit in this triangular system. We must first estimate \( t \) before this step can be made.)

Example. Suppose \( t > 3 \) and

\[ y_3 = \nu_{13} y_1 + \nu_{23} y_2 + \lambda_3 h + U_3. \]
Use \( y_t^* = \lambda_t h + U_t \) as a proxy for \( h \). Substituting this into our \( y_3 \) equation yields:

\[
y_3 = \nu_{13}y_1 + \nu_{23}y_2 + \frac{\lambda_3}{\lambda_t} y_t^* + \left( U_3 - \frac{\lambda_3}{\lambda_t} U_t \right).
\]

Observe that \( y_1 \), and \( y_2 \), are independent of our residual, but \( y_t^* \) is not. We can use \( y_{t\beta}^* \) as an instrument for \( y_{t\alpha}^* \) to estimate the parameters above. This obviously generalizes for all equations less than \( t \).
iv. Assume identification for all equations through $t$ via an exclusion restriction in equation $t$.

**Example.** As an example, consider the following:

$$y_4 = \nu_{14}y_1 + \nu_{24}y_2 + \nu_{34}y_3 + \lambda_4 h + U_4$$

Define:

$$y_2^* \equiv y_2 - \nu_{12}y_1, \quad y_3^* \equiv y_3 - \nu_{13}y_1 - \nu_{23}y_2$$

Solving for $y_1$ and $y_2$ and substituting into the equation for $y_4$, we find:
\[ y_4 = \nu_{14} y_1 + \nu_{24} y_2 + \nu_{34} (y_3^* + \nu_{13} y_1 + \nu_{23} y_2) + \lambda_4 h + U_4 \]

\[ = (\nu_{14} + \nu_{34} \nu_{13}) y_1 + (\nu_{24} + \nu_{34} \nu_{23}) y_2 + \nu_{34} y_3^* + \lambda_4 h + U_4 \]

\[ = (\nu_{14} + \nu_{34} \nu_{13}) y_1 + (\nu_{24} + \nu_{34} \nu_{23})(y_2^* + \nu_{12} y_1) \]

\[ + \nu_{34} y_3^* + \lambda_4 h + U_4 \]

\[ = \nu_{14}^* y_1 + \nu_{24}^* y_2^* + \nu_{34} y_3^* + \lambda_4 h + U_4 \]

where:

\[ \nu_{14}^* = \nu_{14} + \nu_{24}^* \nu_{12} + \nu_{34} \nu_{13} \]

\[ \nu_{24}^* = \nu_{24} + \nu_{23} \nu_{34} \]

Using \( y_1 \) as a proxy for \( h \) and substituting we get:

\[ y_4 = \pi_1 y_1 + \nu_{24}^* y_2^* + \nu_{34} y_3^* + \left( U_4 - \frac{\lambda_4}{\lambda_1} U_1 \right) \]
where $\tau_1 = \nu_{14}^* + \frac{\lambda_4}{\lambda_1}$. We can then use $y_{1\beta}, y_{2\alpha}^*$, and $y_{3\alpha}^*$ as instruments to get an estimate of $\nu_{34}$. Define:

$$\tilde{y}_4 = y_4 - \nu_{34}y_3 = \nu_{14}y_1 + \nu_{24}y_2 + \lambda_4 h + U_4$$

(Excluding $y_3$ allows us to estimate the remaining parameters). Using $y_3^*$ as a proxy for $h$ yields:

$$y_4 = \nu_{14}y_1 + \nu_{24}y_2 + \frac{\lambda_4}{\lambda_3}y_3^* + \left( U_4 - \frac{\lambda_4}{\lambda_3} U_3 \right).$$

We can then estimate $\nu_{14}$, and $\nu_{24}$ by using $y_{1\alpha}, y_{2\alpha}^*$, and $y_{3\beta}^*$ as an IV. We can continue estimating. For example, consider the $5^{th}$ equation:

(i) Rewrite in terms of $y_1, y_2^*, y_3^*$, and $y_4^*$. 


(ii) Use $y_1$ to proxy $h$.

(iii) Use a cross-member $IV$ for $y_1$ in addition to $y_j^*, j = 2, 3, 4$ which gives our estimate of $\nu_{45}$.

(iv) Now form $\tilde{y}_5 = y_5 - \nu_{45} y_4$.

(v) With $y_4$ excluded, we can use purged $IV$’s on $\tilde{y}_5$, as before.
3.4 Comments

1. One needs to check the rank order conditions for identification (requires imposing an exclusion restriction).

2. Griliches and Chamberlain (IER, 1976) find a small ability bias - 3rd decimal point difference in schooling coefficient.
4 Twin Methods

Basic Principle: Monozygotic or MZ (identical) twins are more similar than Dizygotic or DZ (fraternal) twins. The key assumption is that if environmental factors are the same for both types of twins, then we can estimate genetic components to outcomes.
4.1 Univariate Twin Model

Let $y = \text{observed phenotypic variable}$, $x = \text{unobserved genotype}$, and $u = \text{environment}$. Further, suppose that we can write our model additively:

$$y = x + u$$

and assume independence of $x$ and $u$ so that $\sigma_y^2 = \sigma_x^2 + \sigma_u^2$. Now suppose that we have data on another individual:

$$y^* = x^* + u^*$$

Then our phenotypic covariance is:

$$\text{Cov} (y, y^*) = \text{Cov} (x, x^*) + \text{Cov} (u, u^*)$$
where we are imposing the assumption:

\[ \text{Cov} (x, u^*) = \text{Cov}(x^*, u) = 0. \]

Defining standardized forms and some simplifying notation, let

\[ \tilde{y} \equiv \frac{y}{\sigma_y}, \quad \tilde{x} \equiv \frac{x}{\sigma_x}, \quad \tilde{u} \equiv \frac{u}{\sigma_u}, \quad h^2 \equiv \frac{\sigma_x^2}{\sigma_y^2}, \quad \rho^2 \equiv \frac{\sigma_u^2}{\sigma_y^2} \]

Thus, \( \tilde{y}\sigma_y = \tilde{x}\sigma_x + \tilde{u}\sigma_u \) which implies \( \tilde{y} = h\tilde{x} + \rho\tilde{u} \). We can also derive the identity:

\[ h^2 + \rho^2 = \frac{\sigma_x^2}{\sigma_y^2} + \frac{\sigma_u^2}{\sigma_y^2} = 1 \]

where the last step follows from our assumption of independence. Now we wish to consider the correlation between ob-
served phenotypes of our two individuals:

\[ C = Corr(y, y^*) \]
\[ = Corr(h\tilde{x} + p\tilde{u}, h\tilde{x}^* + \rho\tilde{u}^*) \]
\[ = h^2 \frac{Cov(\tilde{x}, \tilde{x}^*)}{Var(\tilde{x})} + \rho^2 \frac{Cov(\tilde{u}, \tilde{u}^*)}{Var(\tilde{u})} \]
\[ = h^2 g + \rho^2 \nu \]

say, with \( g \) and \( \nu \) defined as above. We assume that \( g_{MZ} = 1 \) and that \( g_{DZ} < 1 \). That is, the genotypic variable is perfectly correlated among identical twins, but less than perfectly correlated among fraternal twins. Replacing this result into the above produces:

\[ C_{MZ} = h^2 + \nu_{MZ}\rho^2 \]
\[ C_{DZ} = h^2 g_{DZ} + \nu_{DZ}\rho^2 \]
Therefore:

\[ C_{MZ} - C_{DZ} = (1 - g_{DZ})h^2 + (\nu_{MZ} - \nu_{DZ})\rho^2 = (1 - g_{DZ})h^2 + (\nu_{MZ} - \nu_{DZ})(1 - h^2) \]

where the last equality follows from our established identity. Solving for \( h^2 \), we find:

\[
h^2 = \frac{(C_{MZ} - C_{DZ}) - (\nu_{MZ} - \nu_{DZ})}{(1 - g_{DZ}) - (\nu_{MZ} - \nu_{DZ})}.
\]

The only known in the right hand side of the above equality is the expression \((C_{MZ} - C_{DZ})\), which is simply the correlation coefficient of the observed phenotypic variable. The remaining two expressions, \((1 - g_{DZ})\) and \((\nu_{MZ} - \nu_{DZ})\) can not be computed as they represent statistics on variables we don’t observe.
One could impose $\nu_{MZ} = \nu_{DZ}$ so that:

$$h^2 = \frac{C_{MZ} - C_{DZ}}{1 - g_{DZ}}.$$ 

The expression $g_{DZ}$ is a measure of how closely the genetic variable is correlated across our two observations. One could then guess or estimate a value for this parameter to derive corresponding estimates of $h^2$, the ratio of how much variance in the phenotypic variable is explained by variance in the genetic component. Other studies have attempted to include $Cov(x, u) \neq 0$ but this presents an identification problem. A typical value of the estimable portion of the above, $C_{MZ} - C_{DZ}$, is commonly reported in the literature to be 0.2.
Notes on Figure 2, Predictive Validities of IQ and Big Five Dimensions

Leadership

Associations between personality and IQ and leadership were taken from two meta-analyses conducted by the same research group (Judge, Bono, Ilies, and Gerhardt, 2002; Judge, Colbert, and Ilies, 2004). Leadership was defined jointly as “leader emergence,” the degree to which the individual is viewed as a leader by others, and “leader effectiveness,” performance in influencing and guiding the activities of a group. Typically, these assessments were made by subordinates, supervisors, peers, or observers. Studies relying on self-report assessments of leadership were not included.

Estimated true score correlations between IQ and leadership were corrected for unreliability in the predictor and criterion, as well as for range restriction. Estimated true score correlations between personality and leadership were corrected for reliability in the predictor and criterion, but not for range restriction.

Job Performance

Associations between personality and job performance were taken from a meta-analysis (Barrick and Mount, 1991). Job performance was defined by three criteria: job proficiency (primarily assessed by performance ratings), training proficiency (primarily assessed by training performance ratings), and personnel data (including salary level, turnover, status change, and tenure).

The association between IQ and job performance was taken from (Hogan, 2005). This article did not state whether this correlation is observed or corrected. The much higher estimate of corrected validity is offered by Schmidt and Hunter (2004). Concerns about over-correction with respect to restriction on range and reliability have been raised by Hartigan and Wigdor (1984) with specific reference to estimating the effect of IQ on job performance.

Longevity

Associations between personality and longevity were taken from a review of 34 studies that were prospective in design and which controlled for demographic factors (Roberts, Kuncel, Shiner, Caspi, and Goldberg, in press). Estimated true score correlations were not provided.

Years of Education

Cross-sectional associations between personality and years of schooling were taken from a study (Goldberg, Sweeney, Merenda, and Hughes, 1998) using a large (N = 3629) sample of individuals representative of working adults in the U.S. in the year 2000. Estimated true score correlations were not provided.
The association between IQ and years of schooling was taken from a review by an American Psychological Association (APA) taskforce (Neisser et al., 1996). This article did not state whether this correlation was observed or corrected.

**College Grades**

Associations between personality and college academic performance were taken from a meta-analysis of 23 studies (collective $N = 5878$) by O’Connor and Paunonen (in press). Most of the reviewed studies used as measures of academic performance GPA, but several also used course exam grades.

The association between IQ and college GPA is from a review (Jensen, 1998). This article did not state whether this correlation is observed or corrected. A similar estimate ($r = .5$) is offered by Neisser et al. (1996) for the association between IQ and general academic performance.

**References**


