Comment on the Identification Strategy in "Family Ruptures, Stress, and the Mental Health of the Next Generation", Part 2

Matsumoto, Brett

July 2016

Comment on the Identification Strategy in "Family Ruptures, Stress, and the Mental Health of the Next Generation", Part 2

Brett Matsumoto[†]

Bureau of Labor Statistics[‡]

July 8, 2016

Abstract

The empirical methodology used by Persson and Rossin-Slater (2016) to estimate the causal effect of *in utero* exposure to stress on later outcomes has two sources of bias. First, they define the control group in such a way that biases their causal estimates and can lead to the finding of a significant relationship when there is none. Therefore, the causal effect that the authors claim to establish may in fact be spurious correlation from an error in their empirical methodology. Second, measurement error in the estimation of gestation age can exacerbate the magnitude of the bias caused by the misspecified control group. An additional concern is that the endogeneity of actual birth date may make little practical difference in this context, which undermines the author's claim of being the first to establish a causal effect.

1 Introduction

In their paper "Family Ruptures, Stress, and the Mental Health of the Next Generation", Persson and Rossin-Slater attempt to estimate the causal effect of *in utero* exposure to

[†]Contact: matsumoto.brett@bls.gov. This version primarily focuses on criticisms not contained in the first version of the comment. It should therefore be viewed as an extension of the earlier version, although some portions of the earlier comment are included for the sake of completeness. Prior to writing this comment, I posted an outline of the criticisms contained in this document in various online forums. I can certify that I am the author of the relevant posts and that the ideas put forth here are my own. I would particularly like to acknowledge the support of anonymous economists at econjobrumors.com who played a critical role in helping me to formulate the ideas contained within this note.

[‡]The views and opinions expressed in this paper are my own and do not reflect the views of the BLS, Department of Labor, or the Federal Government.

stress and mental health outcomes later in life. They do this by considering mothers who experience the loss of a close relative during pregnancy as the treatment group. The control group is mothers who experience this loss shortly after birth. Let c denote conception, b denote birth, and e_b denote expected date of birth ($e_b = c + 280$ days). Then the authors estimate the causal effect using the following equation (equation 5 in the paper):

$$Outcome = \beta_0 + \beta_1 \mathbf{1}[c \leq \ RelativeDeath \ < e_b] + \epsilon, \tag{1}$$

where the estimation sample is limited to mothers who experience the relative death during pregnancy or within one year of actual birth. The concern with defining the treatment variable as the relative death occurring prior to the actual date of birth is that the actual date of birth may be correlated with the error term. On the one hand, there is the mechanical relationship that a relative death is more likely to occur during pregnancy the longer the pregnancy lasts. Also, there is the concern that the treatment defined using actual date of birth is endogenous since relative death can affect the length of the pregnancy. The authors claim that defining the treatment group based on the expected date of birth corrects for the endogeneity. Alternatively, the causal parameter could be estimated using treatment based on expected date of birth as an instrument for treatment based on actual date of birth.

This identification strategy suffers from two sources of bias. First, defining the control group using deaths that occur one year from the actual date of birth creates spurious correlation that biases the causal estimates. Second, measurement error in gestation age can bias the results, particularly for outcomes that are defined in terms of actual gestation age. Finally, even if the control group is defined correctly and measurement error is small, endogeneity of actual date of birth may make little practical difference in this context. The next few sections describe these issues in greater detail and provide simulation evidence that the empirical methodology of Persson and Rossin-Slater (2016) is flawed and that their results are biased.[1]

2 Misspecified Control Group

The control group is defined implicitly through the sample restriction and includes mothers who experience the death of a relative between the *expected* date of birth and one year after the *actual* date of birth. The correct specification uses the expected date of birth to define all

[1]Black et al. (2016) use a similar identification strategy but also include mother fixed effects. Some of the issues raised in this comment also apply to Black et al. (2016), but the main concern of bias due to misspecification of the control group does not apply. Also, a variation of this identification strategy was also used by Currie and Rossin-Slater (2013); Rossin-Slater (2013) and some of the issues raised here apply to those papers as well.

of the treatment and control thresholds. By defining the end of the control group as one year after actual birth, the length of the period that defines the control group differs depending on when a child is born relative to the expected birth date. For the treatment effect (β_1) to be unbiased, not only does there have to be no selection into treatment, there also has to not be selection into the control group (since the treatment effect is defined as the conditional (on observables) difference in mean outcome between the control and the treatment group). The probability that someone is control (versus out of the sample) depends on the length of the control group interval (e_b to $b + 365$). Since individuals who are born prematurely ($b < e_b$) have a shorter control group window, these individuals are less likely to be in the control group. For any outcome correlated with actual gestation length, the average outcome of the control group will be skewed towards individuals with high gestation length. Since the average outcome of the control group is biased, the treatment effect is biased in the opposite direction. For example, consider the case where the average gestation length is the same for all groups. The average for the treatment group is equal to the true population average, but the average for the control group will be higher because individuals with longer actual gestation lengths are overrepresented in the estimation sample. From this, one would falsely conclude that the treatment effect on gestation length is negative. Virtually all of the outcomes considered are correlated to some degree with actual gestation length (e.g., birth weight, health conditions at birth or shortly after, and even later life outcomes), which means that the misspecification of the control group causes bias in all of the results. The direction of the bias is in the direction of the estimated results, so the authors cannot conclude that their significant results are due to there actually being a true causal effect rather than just the result of careless empirical work.[2]

3 Measurement Error in Gestation Age

The second concern is related to the use of gestation age to generate the exogenous treatment variable based on the expected date of birth. The expected date of birth is defined as the date of conception plus 280 days. The problem is that the date of conception is not known, so the authors calculate it using the gestation age, which is reported in the data. Gestation age is itself merely an estimate, so the conception date as calculated is measured with error.[3]

[2]See Black et al. (2016) for an example of this identification strategy that uses a correctly specified control group.

[3]The initial estimate of conception date is 14 days after the start of the last menstrual cycle. This estimate is later refined based on prenatal ultrasounds. If the estimated gestational age based on the ultrasound is significantly different from that based on the last menstrual cycle, then the ultrasound estimate is used as the preferred measure (see American College of Obstetricians and Gynecologists (2014) for details). Significantly, the accuracy of the estimate of gestational age based on ultrasound decreases as the pregnancy progresses.

The authors do not provide any detail in their paper about how the gestation age is measured but have informed me that the measure used is based on the last menstrual cycle. The estimate based on last menstrual cycle will have greater measurement error than alternative measures, but the magnitude of the error will be equal for the treatment and control groups. There is a problem when this identification strategy is used for outcomes that are defined based on gestation age. This is because the error in measuring gestation age will occur in both the LHS outcome variable and in the RHS definition of the treatment variable. Measurement error in the LHS variable will be captured in the error term. Therefore, the measurement error in gestation age will cause the error term to be correlated with the treatment variable and will bias the estimate of the causal effect of the treatment. With a properly specified control group, the measurement error does not appear to create much bias in the results.[4] When the control group is misspecified, however, the measurement error will increase the bias.[5] The amount of additional bias depends on the variance of the measurement error. I provide simulation evidence that the effect of both the misspecified control group and measurement error in gestation age cause bias in the empirical specification of Persson and Rossin-Slater (2016). Specifically, the simulation evidence shows that one can obtain a statistically significant estimate of the treatment effect in the direction of the results reported in their paper when the true effect is zero.

4 Endogeneity of Actual Date of Birth

An additional concern with the identification strategy is that the endogeneity of defining treatment using the actual date of birth may make little practical difference. In Persson and Rossin-Slater (2016) the endogeneity of actual date of birth may not generate much bias in the OLS estimate based on the results of the first stage regression (R^2 of approximately

The most accurate estimate of gestational age is obtained from an ultrasound at between 10-12 weeks. By the 24 week mark, the estimate based on the ultrasound is no more precise than that based on last menstrual cycle (Verburg et al., 2008). The estimate of gestation age based on ultrasound will be more precise, but the treatment may affect how individuals time their prenatal care. In order to use the estimate of gestation age based on ultrasound, researchers will have to show that the treatment is unlikely to affect the estimate of gestation age as individuals in the treatment group may delay or skip prenatal appointments if they coincide with the treatment (e.g., death of a close relative, hurricane, etc.). Particularly if these missed appointments are early in the pregnancy, the treatment group may have poorly estimated gestation ages relative to the control group.

[4]This finding is based on the results of a few relatively simple simulations. Anyone wishing to use this identification strategy should perform additional analysis to determine if measurement error with a properly specified control group is likely to affect their results.

[5]This is because the measurement error in gestation age affects all of the thresholds equally, except the endpoint of the control group. Since the measurement error affects the control group differently than the treatment group, the estimate of the treatment effect (which is the difference) will be biased.

0.97, Table D-1) and the statement that "the instrument (relative death before expected birth date) is different from the actual exposure variable (relative death before actual birth date) for only about 1 percent of the individuals in our data."

In a related paper Black et al. (2016), present the results using the same identification strategy on different data and outcomes. They report that the first stage R^2 is very close to one and that the OLS estimates and the 2SLS estimates are very similar when considering later in life outcomes, but they do find differences when considering birth outcomes.[6] None of the studies that use expected date of birth to instrument for actual date of birth conduct statistical tests for endogeneity of actual date of birth for the various outcomes.

The problem for the authors is that they differentiate their results from a vast prior literature that does not control for this endogeneity. Since they do not perform any statistical tests for endogeneity and do not even report the OLS coefficients using actual date of birth to define the treatment, it is impossible for the reader to evaluate whether endogeneity is a significant problem in this context. If endogeneity is not a problem, then instead of being the first to estimate the causal relationship, the authors are merely verifying that prior estimates in the literature have a causal interpretation.

5 Simulation Results

The purpose of this simulation is to show that the issue of a misspecified control group and measurement error in gestation age can cause the identification strategy of Persson and Rossin-Slater (2016) to falsely attribute a significant causal effect to death *in utero*. Conception occurs for all individuals at a given period C. The average pregnancy lasts 40 periods.[7] Let ϕ denote a draw from a standard normal distribution, then birth occurs at $B = C+40+\sigma_B\phi$, where σ_B^2 is the variance of the distribution of births. True gestation age at birth is $GA = B-C$. The estimate of gestation age at birth is given by $\hat{GA} = B-C+\sigma_{GA}\phi$, where the accuracy of the estimate is determined by σ_{GA}. Then, the estimated date of conception is $\hat{C} = B - \hat{GA}$.

Deaths occur once for each individual and are distributed uniformly over the interval 0 to T. The treatment group is defined as the death occurring in the interval between $\hat{C}$ and the predicted date of birth, $\hat{C}+40$. The control group is implicitly defined as the death occurring between B and $B+50$ by limiting the estimation sample to anyone who experiences a death

[6] Other studies, such as Currie and Rossin-Slater (2013) and Rossin-Slater (2013), in different contexts but still looking at *in utero* effects of some treatment find that instrumenting for treatment based on actual date of birth using expected date of birth does yield different results. Those studies instrument for not only the endogeneity of pregnancy length but also location change. This yields a lower first stage R^2.

[7] A period can be thought of corresponding to one week.

between $\hat{C}$ and $B + 50$. The simulations consider 4 different outcomes. The first two are gestation age at birth and the probability of preterm birth (gestation age at birth less than 37). These outcomes have the gestation age measurement error in both the RHS definition of the treatment group and the LHS outcome variable. The last 2 outcomes are correlated with actual gestation age and differ by the amount of correlation. One outcome (Out 1) is strongly correlated with actual gestation age and can be thought of as birth outcomes such as birth weight, health conditions at birth, or birth complications. The other outcome (Out 2) is correlated with the first causing it to be more weakly correlated with actual gestation age. Since birth weight is associated with many later in life outcomes, this second outcome can be thought of corresponding to later in life outcomes.

The only explanatory variables are an intercept term and an indicator for treatment (death between $\hat{C}$ and $\hat{C} + 40$). For comparison, the results are also shown defining the treatment using the actual date of conception C. These results allow the bias to be decomposed between the spurious correlation caused by measurement error and the spurious correlation caused by the misspecified control group. The simulations differ in the variance of the error of the estimate of gestation age and the sample size with 1000 iterations used for each combination of these parameters. The actual gestation age is used in the case where the variance of the error in gestation age equals 0.

The following values are used in the simulation: $C = 20$, $\sigma_B = 2$, $T = 130$, $Out1 = .2 * GA + \phi$, $Out2 = -.25 * Out1 + \phi$, where each ϕ is a draw from a standard normal distribution.

As the sample size increases, the standard deviation of the estimates fall while the point estimates are roughly the same. Therefore, the result of increasing the sample size is an increase in the statistical significance of the treatment on all of the outcomes. As the error in the measure of gestation age increases, the bias in estimating the effect on gestation age related outcomes increases. This implies that noisier estimates of gestation age such as date of last menstrual cycle will be more biased than estimates based on more accurate methods such as ultrasound. Another important fact is that the direction of the bias is negative in terms of total gestation length. Therefore, the significant negative findings in the literature of a significant negative effect of *in utero* relative death on total gestation length may be due to measurement error rather than reflecting a true causal relationship. Additionally, even if gestation age was measured perfectly, the estimates are still biased based on how the treatment and control groups are defined. The differences in the results of Persson and Rossin-Slater (2016) and Black et al. (2016) are due, at least in part, to differences in how the control groups are defined.

For the outcomes that are correlated with gestation age at birth, the bias does not

Table 1 : Simulation Results

			σ_{GA}		
Outcome	0.0	0.5	1.0	1.5	2.0
N=10,000					
GA	-0.082 (0.049)	-0.087 (0.051)	-0.102 (0.055)	-0.127 (0.058)	-0.162 (0.065)
Pre-term	0.005 (0.006)	0.006 (0.006)	0.007 (0.006)	0.010 (0.007)	0.013 (0.009)
Out 1	-0.017 (0.026)	-0.017 (0.026)	-0.017 (0.026)	-0.017 (0.026)	-0.017 (0.026)
Out 2	0.005 (0.009)	0.005 (0.009)	0.005 (0.009)	0.004 (0.010)	0.004 (0.010)
N=100,000					
GA	-0.082 (0.015)	-0.085 (0.016)	-0.101 (0.017)	-0.126 (0.019)	-0.160 (0.021)
Pre-term	0.005 (0.002)	0.006 (0.002)	0.007 (0.002)	0.010 (0.002)	0.013 (0.003)
Out 1	-0.016 (0.009)	-0.016 (0.009)	-0.016 (0.009)	-0.016 (0.009)	-0.016 (0.009)
Out 2	0.004 (0.003)	0.004 (0.003)	0.004 (0.003)	0.004 (0.003)	0.004 (0.003)
N=1,000,000					
GA	-0.080 (0.005)	-0.085 (0.005)	-0.100 (0.005)	-0.125 (0.006)	-0.160 (0.007)
Pre-term	0.005 (0.001)	0.006 (0.001)	0.007 (0.001)	0.010 (0.001)	0.013 (0.001)
Out 1	-0.016 (0.003)	-0.016 (0.003)	-0.016 (0.003)	-0.016 (0.003)	-0.016 (0.003)
Out 2	0.004 (0.001)	0.004 (0.001)	0.004 (0.001)	0.004 (0.001)	0.004 (0.001)

Notes: Standard deviation in parenthesis. Average over 1,000 iterations.

appear to depend on how accurately gestation ages are measured. The source of the bias is essentially entirely due to the definition of the control group. The next set of simulations further examines this source of bias by looking at $Out1 = GA * .2 + \rho * \phi$. For these simulations, the variance in the gestation age measure is set at 1 and the correlation of Out 1 with actual gestation age is varied by changing ρ. The magnitude of the bias does not change very much as the sample size increases or as the outcome becomes less correlated with the actual gestation age. The significance of the estimated effect decreases as the variance of the independent variation in the outcome increases. The bottom panel of the table shows the results when the predicted date of birth is used to define the control group (death within one year of the predicted date of birth). For these results the bias disappears.

These simulations show defining the control group using one year from the actual date of birth causes bias for any outcome that is correlated with the actual gestation age at birth. Measurement error of gestation age causes additional bias for outcomes that are defined in terms of gestation age at birth. The misspecification of the control group creates spurious correlation that bias nearly all of the estimates reported by Persson and Rossin-Slater (2016). As long as this source of bias remains, there is no way to determine if the statistically significant relationships they find are true causal relationships or merely the result of their erroneous empirical specification.

Table 2 : Additional Simulation Results

	ρ			
Outcome	0.1	0.5	1.0	2.0
N=10,000				
Out 1	-0.016 (0.010)	-0.016 (0.016)	-0.016 (0.026)	-0.018 (0.049)
N=100,000				
Out 1	-0.016 (0.003)	-0.016 (0.005)	-0.016 (0.008)	-0.016 (0.015)
N=1,000,000				
Out 1	-0.016 (0.001)	-0.016 (0.002)	-0.016 (0.003)	-0.016 (0.005)
Treatment defined using predicted birth				
N=100,000				
Out 1	1.0e-4 (0.003)	1.0e-4 (0.005)	0.000 (.008)	7.0e-4 (0.016)

Note: Standard deviation in parenthesis. Average over 1000 iterations.

6 Conclusion

In light of the evidence presented in this comment, the authors and future researchers should take additional steps to address these issues. For researchers who wish to use this identification strategy in their own work I recommend (in addition to specifying the control group correctly) that they report more detail on how gestation age is measured and consider the effect that measurement error may have on their results. Additionally, given that treatment based on expected date of birth almost perfectly coincides with treatment based on actual date of birth, researchers should conduct statistical tests for endogeneity. At a minimum, researchers should report the results for treatment defined both ways.

Persson and Rossin-Slater (2016) claim as their primary contribution that they are the first to estimate a causal relationship between *in utero* exposure to stress and later in life mental health outcomes. There are two problems with this claim. First, if endogeneity of actual date of birth is not important when considering later in life mental health outcomes, then the correlations from the prior medical literature have a causal interpretation. Therefore, the authors are not the first to estimate a causal relationship. A simple statistical test for endogeneity could resolve this issue. The second problem is that the estimates from their empirical specification contain bias from their incorrectly specified control group, which is compounded by measurement error in their estimate of gestation age. Luckily for the authors, this issue is also easily addressable. All they need to do is to rerun their analysis with the correctly specified control group. The authors have informed me that their article is final and refuse to perform this very simple additional analysis. They are content to leave

in doubt the primary contribution of their paper. What ever happened to taking pride in ones work?

References

Black, S., P. Devereux, and K. Salvanes (2016). Does grief transfer across generations? bereavements during pregnancy and child outcomes. *American Economic Journal: Applied Economics 8*(1).

Currie, J. and M. Rossin-Slater (2013). Weathering the storm: Hurricanes and birth outcomes. *Journal of Health Economics 32*(3), 487–503.

Persson, P. and M. Rossin-Slater (2016). Family ruptures, stress, and the mental health of the next generation. *American Economic Review forthcoming*.

American College of Obstetricians and Gynecologists (2014). Committee opinion no. 611: Method for estimating due date. *Obstetricics and Gynecology 124*, 863–866.

Rossin-Slater, M. (2013). Wic in your neighborhood: New evidence on the impacts of geographic access to clinics. *Journal of Public Economics 102*(0), 51–69.

Verburg, B. O., E. A. P. Steegers, M. D. Ridder, R. J. M. Snijders, E. Smith, A. Hofman, H. A. Moll, V. W. V. Jaddoe, and J. C. M. Witteman (2008). New charts for ultrasound dating of pregnancy and assessment of fetal growth: longitudinal data from a population-based cohort study. *Ultrasound in Obstetrics & Gynecology 31*(4), 388–396.