**A Comment on Dench and Joyce (2019)**
***The Earned Income Tax Credit and Infant Health Revisited***

**Hilary Hoynes, Douglas Miller and David Simon**
**November 22, 2019**

Dench and Joyce in a manuscript circulating since summer 2018, provide a comment on our paper published in *AEJ Policy* in 2015 (Full citation: Hilary Hoynes, Doug Miller and David Simon. "Income, the Earned Income Tax Credit, and Infant Health," *American Economic Journal: Economic Policy*, Vol. 7, No. 1, February 2015, pp. 172-211).

This is the third version of the manuscript that we have read and responded to. The specific arguments made and results reported have shifted in each subsequent draft. For those interested, you can find our responses to the earlier versions of Dench and Joyce's manuscript here and here.

Here, we discuss Dench and Joyce's November 2019 paper.

In this Comment we refer to the current manuscript as DJ and we refer to our 2015 *AEJ: Economic Policy* paper as HMS.

Overview of identification strategy in HMS

HMS use variation in the federal EITC to estimate the effects of the credit on infant health (low birth weight). We use two identification strategies. The main identification strategy is to leverage the variation from the 1993 expansion of the federal credit, which varied by parity (of birth) and tax year. HMS use difference-in-difference and event study models applied to data for 1991-1998 – in both models forming treatment and control groups using parity of birth. In one model we compare births of parity 2+ to births of parity 1 (we refer to this as DD1). In another model we simultaneously compare births of parity 3+ to parity 1 and parity 2 to parity 1 (we refer to this as DD2). In the third model, we compare parity 3+ to parity 2 (we refer to this as DD3). In the second identification strategy we expand the time frame to encompass the 1984, 1990, and 1993 EITC expansions. We estimate a panel fixed effects model where we measure the generosity of the EITC using the maximum EITC credit.

In their critique of HMS, DJ make 5 main points:

- Comparing 1$^{st}$ births to 2$^{nd}$ births for OBRA 1993 difference-in-difference and event study analysis is problematic, because the earlier 1990 EITC expansion for 2$^{nd}$ births phased in over the "pre-period" for the 1993 expansion.
- Comparing 2$^{nd}$ births to 3$^{rd}$ and higher births for OBRA 1993 difference-in-difference and event study analysis is problematic because a placebo test fails (4+parity vs. parity 3, for African American mothers).
- The identified mechanisms are insufficient to explain the birthweight impacts
- There is difficulty in identifying effect of a national policy at a single point in time
- There is sensitivity to adding quadratic trends by parity to panel fixed effects model using multiple expansions

As we review these critiques we conclude that many of them are not new and are already discussed in HMS, or are not relevant as they critique models or methods not done in HMS. Others, we argue, are unfounded. Finally, some of the critiques are novel and founded. We distinguish between these different types of critiques below as we review the main points in DJ.

**<u>DJ Point 1:</u> Comparing 1st births to 2nd births for OBRA 1993 difference-in-difference and event study analysis is problematic (presented in DJ Section I).**

A central criticism in DJ is that when comparing to parity 1 births HMS use a pre-period of 1991-1993 to estimate the OBRA 1993 difference-in-differences and event study models. For example on page 5 DJ state:

A. *"These staggered expansions make comparisons between women of parity 2 versus parity 1 almost impossible to identify when the study period is limited to effective tax years 1991-1998 as used by HMS."*

First, it is fundamental to note that DJ refer here to a model that compares parity 2 versus parity 1 as a research design. **HMS never estimated such a model; thus a large portion of DJ's comment is centered on a criticism of a model that is not in HMS**. HMS always pools parities: our main models compare parities 2+ vs. 1, compares parities 3+ vs. 2, or estimates a pooled model comparing 3+ vs 1 and 2 vs. 1.

Regardless, our choice of 1991-1993 for the pre-period was a deliberate, pre-determined choice, based on the timing of the tax change(s) and following the practice in the existing literature.

As HMS show in Figure 1, OBRA93 is the largest expansion of the EITC and also the only one where the policy expands differentially across three parity groups (1, 2, 3+ corresponding to childless, one child, and two or more children). HMS Figure 1 shows that the OBRA93 expansion took place between 1993 and 1994 for parity 1 and parity 2. For parity 3 and higher, the tax change was phased in between 1994 and 1996. HMS Figure 1 also shows a previous tax change (OBRA90) that expanded the credit for parity 2 or more.[1] As HMS show, the vast majority of the OBRA90 change (impacting parities 2+) occurred by 1991 (with smaller increases in 1992, 1993). This is a key factor motivating HMS's decision to use 1991-1993 as the pre-period for analyzing OBRA93 when looking at 2+ vs. 1 births (rather than 1990-1993).

We illustrate this in the first panel in Table 1 below, which shows the maximum benefits by family size (parity) and year in panel 1, the difference (relative to 1993) in the *maximum credit* for each parity in panel 2, and the difference-in-difference (relative to 1993, relative to parity 1) in panel 3. We also show the difference-in-difference below in Figure 1.

Second, our choice to use OBRA93 in a difference-in-difference model and with a (short) pre-period follows other papers in this literature. For example, Evans and Garthwaite (2014) use a similar approach to examine impacts on maternal health. There is a large literature examining the effects of the EITC on employment of single mothers. Many papers follow the approach used in HMS. Typically, one starts with the comparison of women with kids to women without kids (parity 2+ vs parity 1). Some papers then go on to limit the analysis to women with kids and compare those with 2+ children compared to 1 child (parity 3+ vs parity 2). (For a review of this literature see Nichols and Rothstein 2016.)

[1] Childless filers, those of parity 1, have no EITC until 1994.

A second criticism DJ raise in this section is:

B. *If you use 1990-1993 as a pre-period there is a large increase in the EITC in the pre period that invalidates the design*

When DJ claim that the pre-period increase for parity 2 vs parity 1 exceeded the post period difference **(“***From 1990 to 1993 the EITC credit increased by $401 ($222 + $179 in 1995 $) for women of parity 2 relative to parity 1 which exceeded the post-period increase of $264 ($578 - $314 in 1995 $) from 1994-1998.” DJ page 4***)** they include the increase in the maximum credit between 1990 and 1991 from the first year of the OBRA90 expansion. **HMS never estimate a model with 1990 in the pre-period; thus the DJ critique is centered on model that is not in HMS.** As discussed above, HMS’s starting date of 1991 was a deliberate decision, following the literature, and in recognition of the timing of the 1990 policy change.

Specifically, looking at the first panel in Table 1 shows that the gain in the maximum credit for those with 2 or more children rose by $199 between 1991 and 1993. However, between 1990 and 1991 it rose by $300. There is a similar pattern for those with 1 child. It is this large increase between 1990 and 1991 that led HMS to set the pre-period where they did.[2] (All of these amounts are in 1999 dollars, as reported in HMS.)

DJ note that some of the expansion of OBRA90 for Parity 2+ occurs during years 1992 and 1993, the “pre period” years for HMS main analysis of OBRA93. **This is not new; it is already discussed in HMS:** HMS Figure 1 shows this moderate expansion in these years, and also that the subsequent expansion 1993-1994 is much larger in magnitude.

DJ claim that this invalidates a comparison of parity 2 vs. parity 1. This claim may reflect a misunderstanding of the research strategies in HMS. First, it is not uncommon to implement a DD analysis with “large growth” as the “treatment” of interest. In this way, parity 1 births experience no growth in the EITC in the (1991-1993) pre-period and a small increase in the post period. Parity 2+ births experience small growth in the pre-period and very large growth in EITC in the post period. In fact, HMS directly leverage the difference-in-difference magnitude of the EITC expansion to scale the treatment effects (see HMS Table 4 and discussion therein). HMS also show the changes in the maximum credit on the event study graphs (e.g. see HMS Figure 3B) to illustrate the year by year changes in the credit compared to the year by year changes in birth weight outcomes. This HMS analysis shows that smaller legislated increases in the EITC lead to smaller change in the LBW rate.

**Overall, the DJ critique about a partial expansion of the EITC in the pre-period is unfounded.**

A third criticism DJ raise in this section is:

C. *Challenge of translating EITC tax year to when women were exposed to it*

---

[2] Table 1 and Figure 1 show that the OBRA90 expansions were quite similar for parity 2 and parity 3+ while the 1993 expansion was much larger for parity 3+. This illustrates why HMS present extensions of the “DD3” specification (the comparison of parity 3+ to parity 2) and use a pre-period back to 1987 (HMS Fig 3b).

On page 4 DJ say: "*…limiting the pre-period to 1991-1993 does little to mitigate the cumulative exposure of their treatment group to increases in the EITC during their pre-period. The reason is because HMS's mapping from the child's birth date to the year the mother is exposed to the EITC … is an approximation at best.*"

In HMS Section III we discuss our assumptions that underlie our mapping of birth date to effective tax year.

Our first assumption which we refer to as "cash in hand" assumes that the EITC's impact on infant health "runs through the cash available to the family, which arrives with receipt of the tax refund" (HMS p. 180). HMS provide data (HMS Figure 2) showing that most EITC tax refunds are received in February (55%) and March (another 25%).

Our second assumption is that the "sensitive developmental stage is three months prior to birth. This is motivated by evidence that the third trimester of pregnancy is important for birth weight production" (HMS p. 180). HMS (pp190, fn 14) justify their assumption that the third trimester is the sensitive period for birthweight as follows:

> "For example, the cohort exposed to the Dutch Famine in the third trimester had lower average birth weight than cohorts exposed earlier in pregnancy (Painter, Roseboom, and Bleker 2005). In addition, Almond, Hoynes, and Schanzenbach (2011) show that the impact of exposure to the food stamp program is greatest in the third trimester. Also see the review in Rush, Stein, and Susser (1980)."

These two assumptions form the basis of how HMS translate a birth in a given month-year to the "effective tax year" that the birth is exposed to. As we discuss in HMS Section III, "The assumptions behind this mapping are unlikely to be precisely accurate. However, our identification strategy does not rely on high-frequency time variation."

Additionally, HMS test the robustness to these assumptions. In the first subsection of the robustness section (HMS section VII.A) they discuss three modifications to these two assumptions: (1) varying the timing of the "sensitive period," (2) changing the assumption about when the EITC is received/spent, and (3) incorporating the timing of income coming from labor supply increases induced by the EITC (which form part of the treatment). They present these results in HMS Table 9 and HMS Appendix Table 4. HMS show that the results are robust to these sensitivity tests. **Thus, this DJ critique is not new -- HMS discuss and explore this criticism in the original paper.**

The above critiques lead DJ to advocate for an alternative pre-period:

D. *Years 1987-1990 provide a superior pre period*

DJ re-estimate the event study models using 1987-1990 as the pre-period (DJ Figures 2B, 2D), in part justified by their view that "*1987-1990 provide(s) a superior pre period with 1991-1998 as the more appropriate post period*." (DJ p. 5). They then show that for this alternative analysis of their design, tests of the parallel trends assumption would fail.

We see this analysis as **not very relevant for HMS**. While there are differential trends by parity over 1987-1990 for some subgroups, this is a period excluded by HMS from their analysis (see more on this below in our response to critique E). Additionally, to reiterate the above, DJ's event study compares parity 2 to parity 1 which is never advanced by HMS as a standalone design.

In difference-in-differences designs there is currently no convention about how far back pre-trends should be tested. The further back one looks, the more likely it is that the parallel-trends assumption will fail. In general, our view is that pre-periods near the intervention are more relevant than pre-periods many years in the past for assessing the validity of a research design. Notably, HMS already provided long term trends by parity for context in HMS (Appendix Figure 1), as is discussed more below.

More generally, DJ argue that the longer term trends in birth weight across parities is inconsistent with the EITC:

*E. Trends in parity for earlier years (DJ Figure 2B, 2D)*

DJ present the trends in birth weight by parity of birth and point out that there are differential pre-trends in women of low birth weight between 1987-1989[3]. **This is not new; it is already discussed in HMS:** HMS discuss the longer term trends in birth weight – see HMS Appendix Figure 1, and discussion in HMS p. 187-188.

HMS begin the paper by presenting the difference-in-difference results (HMS Table 2) along with the event study version of these results (HMS Figure 3). After presenting these main DD results, HMS discuss their results within the context of longer-term trends:

> "We provide additional context for these results in online Appendix Figure 1 panels A–D. In these figures we show raw trends in the rates of low birth weight by parity, over the period 1981–1999. In each figure we show low birth weight probabilities relative to the 1993 level. Online Appendix Figure 1, panel A has results for our high-impact group, and shows two main findings. First, the changes we observe in our experimental period occur within long-run decreases in rates of low birth weight for this population (although in the pre-1991 period there are many other policies changing, including earlier expansions of the EITC). Second, the raw trends for 1991–1998 show the same pattern as in our main event study results in Figure 3, panels A–C; thus, these results are not strongly impacted by the covariate controls in the model. In contrast, the trend for all births is one of increasing rates of low birth weight (online Appendix Figure 1, panel B). The long-run trend for white high-impact mothers (online Appendix Figure 1, panel C) is similar to that for all high-impact mothers, while for high-impact black mothers (online Appendix Figure 1, panel D), the trends are less monotonic. All the figures show that there are large changes in parity gaps in low birth weight that occur during the 1980s. It would be valuable, but outside of the scope of this study, to understand these changes better." [HMS p. 187-188]

[3] DJ Figure 2D presents an event study comparing 2+ vs. 1 which shows a decline in low birthweight after 1990. This is consistent with the OBRA90 expansion differentially treating 2+ vs. 1 births, as DJ had already pointed out. The decline in low birthweight births at this time should be taken as evidence that supports our hypothesis that the EITC improved infant health.

In summary, in response to DJ section 1, we believe that the analysis either (1) reflects a recapitulation of results already reported in HMS, (2) reflects a critique of a hypothetical analysis not performed in HMS, or (3) suggests criticism that do not have merit. HMS's choice of the pre period is the direct result of the tax law changes and this approach is commonly used in the EITC literature. Additionally, HMS are clear in pointing out how this period fit into longer term trends. We remain of the view that explaining these longer run trends is interesting, but outside the scope of HMS's paper.

**<u>DJ Point 2:</u> Comparing 2nd births to 3rd and higher births for OBRA 1993 difference-in-difference and event study analysis is problematic because a placebo test fails (4+parity vs. parity 3, for African American mothers) (*DJ section II)***

A second central criticism in DJ is directed in HMS's difference-in-difference analysis of 3+ parity versus parity 2 in the OBRA93 expansion (HMS call this "DD3"). As discussed above, the advantage of this approach is that the OBRA93 expansion was much larger for these larger families while earlier EITC expansions were common for families of parity two and above.

DJ critique this approach as it fails a placebo test:

*A. "There should be no differential effect of the EITC on women of parity 3 versus those of parity 4 or 4+ as they all experience the same increase in the available tax credits" (DJ p. 8).*

In Section II, DJ provides a placebo test of the 3+ vs 2 design in HMS and show that this comparison produces statistically significant results (DJ Table 2). In particular they provide difference-in-difference estimates of parity 4 versus 3 and parity 4+ vs 3. **This is not new; it is already discussed in HMS.**

HMS provide this placebo test in a robustness and sensitivity test section. HMS describe it here:

> "To examine this further, we examine a series of "pair-wise" comparisons of different parity births. Some of these comparisons (e.g., 2 versus 1, 3 versus 2) embed a treatment and some (e.g., 4 versus 3) form a "placebo test" for our estimation method.
>
> We present results in online Appendix Table 6 for the high-impact sample. In the first row we compare second births (treated under the "one child" EITC schedule) to first births (untreated), and so on. The first two rows reinforce our main findings—there was a relative improvement in low birth weight for second births compared to first births, and also for third births compared to second births. The remaining rows of the table compare pairs of birth parities that are both "treated," and we expect to find no estimated effect for these comparisons. This appears to be true for fifth versus fourth and sixth and higher versus fifth. However, we do find that low birth weight improved more for fourth births than for third births, which is not consistent with our expectation. To investigate this finding further, we estimated an event study model for this comparison. This analysis indicates that the 4 versus 3 difference begins in 1995 and grows after that.
>
> The gap between fourth and third births does raise a cautionary note about potential parity-specific trends in birth weight, and our analysis should be interpreted in light of this caution. We believe that despite this, the preponderance of evidence indicates that the EITC does improve child health. First, the timing of these spurious trends does not correspond cleanly with the

> policy change. And second, in our "maxcredit" models, results are robust to inclusion of parity-specific trends." (HMS p. 205)"

DJ also critique the HMS comparison of parity 3+ compared to parity 2 by pointing to differences in impacts across subgroups:

*B. "The results for all women, however, are driven by birth outcomes among black women" (DJ p.8)*

DJ present results by race and ethnicity subgroups (DJ Table 2) and go on to say "*Importantly, there is no association between the EITC and rates of low birth weight among white or Hispanic women*" (DJ p. 8).

However, this is not new**; differences by race were already discussed in HMS.** HMS Table 3 shows that the point estimates are largest for black women[4]:

> "Table 3 shows heterogeneity in effects by race and Hispanic origin within the high-impact sample. The EITC reduced the likelihood of having a low birth weight birth for black mothers of 0.73 percentage points' (relative to a mean of 14.4 percent), more than four times higher than the effect on white mothers (0.13 percentage point decline relative to a mean of 8.1 percent). Interestingly, smaller treatment effects are experienced by Hispanic mothers than by non-Hispanic mothers (−0.13 versus −0.41 in the second+/first parity model). " (HMS p. 188)

HMS go on to estimate a first stage effect of the EITC expansion on income to add interpretation to the results by racial subgroup.

> "As shown in Table 3, the effects of the EITC expansion on LBW is larger for black, single, low-educated women, compared to white, single, low-educated women. Interestingly the CPS analysis, as shown in panels B and C of Appendix Table 1, shows a smaller first-stage effect on after-tax income for black women (compared to white). Thus, the estimated treatment on the treated results are larger for black compared to white women. For example, the comparison of second and higher births compared to first births shows a 5.3 percent reduction for blacks and a 1.1 percent reduction for whites (percent impact of a $1,000 TOT)." (HMS pp 191-192)

The final point in the critique in DJ Section II puts these points together:

*C. The placebo test for parity 4+ versus 3 fails only for the subgroup of black women*

In DJ Table 2, they provide estimates for the placebo tests (e.g. 4 versus 3, 4+ versus 3) for each of the subgroups (white, black, Hispanic) and show that the placebo test fails only for black mothers.

This is novel evidence– what is new is DJ's presentation of the combination of the placebo analysis with the black subgroup analysis. This new evidence helps to provide additional context to the results already presented in HMS of the overall 4+ vs. 3 trends. However, as with the placebo test on all mothers presented in HMS, the decline in low birthweight occurs in 1995 rather than being concurrent with the

[4] We disagree with the interpretation in DJ that there is strong evidence against effects of the EITC on low birthweight births for white or Hispanic mothers. The estimated effects are smaller for these groups, and statistically insignificant. However, the confidence intervals include moderately sized effects.

OBRA93 expansion. While this new evidence is interesting, we remain skeptical that it is enough to overturn our main findings.

**<u>DJ Point 3:</u> The identified mechanisms are insufficient to explain the birthweight impacts (DJ Section III)**

In section III, DJ argue that there are no plausible mechanisms by which the EITC could improve birth outcomes. There are two sub-points in this DJ critique, one about prenatal care and the other about smoking.

We begin with the observation that in HMS we do not undertake a formal mediation analysis, nor do we undertake a definitive accounting of the mechanisms by which the EITC impacts birthweight. Instead, we examine several candidate mechanisms to point toward potential channels of impact. As we note in HMS (p. 200):

> "This discussion of channels of impact is speculative on our part, and our research design does not let us distinguish it from alternative plausible channels. For example, it may be the case that employment itself leads to reductions in smoking, or instead that early prenatal care leads to reductions in low birth weight. Further, it may be that the increases in income lead to reductions in birth weight through other channels, perhaps including better nutrition."

As such we view the set of critiques raised in this part of DJ as essentially **not new; it was already discussed in HMS.** However, we also note an implicit claim in DJ that evidence against one or another particular mechanism is strong evidence against the main findings in HMS. We view this implicit claim **as unfounded**.

A. *Prenatal care is an implausible explanation – they are too small to be consequential*

We agree with DJ that the magnitude of estimated effects on number of prenatal visits is quantitatively small. This critique is **not new; it is directly noted and addressed in HMS** (p. 198).

HMS also estimate impacts on prenatal care quality (Kessner Index) which are reasonably large. They find the quality of having inadequate prenatal care declines by 0.88 pp to 1.105 pp (DD1). These effects are 5% to 9% of the mean. These findings are also consistent with the impact of the EITC moving mothers from Medicaid to private health insurance coverage (not discussed as a potential mechanism in DJ). In conclusion, we continue to believe that there is evidence pointing toward a role of improvements in prenatal care quality.

*B. Smoking effects are implausible – they are too large*

HMS discuss this point in the paper (see HMS, pp. 177-178). They make the point that behaviors such as smoking and drinking could change due to increases in after tax income (e.g. smoking is a normal good). But they additionally point out that increases in employment (a robust finding for the EITC for single mothers) could independently lead to a decrease in smoking (HMS p.178). In the end, changes in income alone are not the only channel for the effect of the EITC on smoking. Thus, as is discussed in HMS, the approach does not presume or rely on smoking being an inferior good.

There are three papers that look at the effects of the EITC on smoking, and find results consistent with HMS. Averett and Wang (2013) use several difference-in-differences designs to show that smoking declines among low education mothers after the OBRA93 expansion. Cowan and Tefft (2012), similarly use difference-in-differences around the OBRA93 expansion to show that smoking declined among unmarried women with less than a college degree. In contrast Kenkel, Schmeiser, and Urban (2014) leverage changes to the maximum state and federal credit over time to show that smoking increases among low income adults. While no paper we know of has resolved the differences between these studies, the pattern of findings fits our hypothesis: papers focusing on mothers or unmarried women find declines in smoking consistent with increases in employment among these groups. Kenkel, Schmeiser, and Urban find increases in smoking when they focus on the population of adults, potentially driven by men who do not change their labor supply in response to the EITC (Kenkel, Schmeiser, and Urban do not present separate results for men and women smokers).

It is also important to recognize that pregnant women have a unique relationship with smoking due to the role that information plays. Elevating women out of poverty could change their information set, the quality of prenatal care they receive, or their decision to act on medical advice related to smoking. These changes could occur in a variety of ways that are not yet understood by the literature and that we are unable to measure, but suggest a more nuanced story than one solely reliant on the income elasticity of smoking. Ultimately, HMS report effects on smoking because that is the evidence provided by our research design. They acknowledge that it is not yet well understood how income, employment, and prenatal smoking interact.

More generally, DJ argue that we don't know much about what might cause low birth weight (and thus don't have sufficient guidance about the possible mechanisms, possibly beyond smoking):

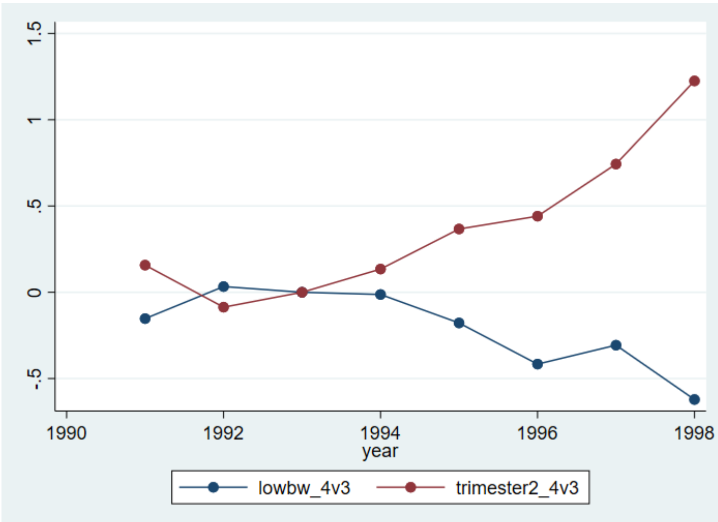
*C. DJ cite a 2007 review by the Institute of Medicine (IOM) which states in the abstract: "The current methods for the diagnosis and treatment of preterm labor are currently based on an inadequate literature, and little is known about how preterm birth can be prevented" (DJ p. 9)*

We agree that there is a lack of understanding of biological mechanisms for individual cases of preterm birth, this is reflected in a low R^2 in many models of preterm birth. However, that does not mean that there are not plausible interventions that can change individual behavioral and environmental factors that influence the occurrence of pre-term birth. This is also recognized by the medical community. The same abstract cited by DJ from the IOM states directly before their quotation: "Preterm birth is a complex cluster of problems with a set of overlapping factors of influence. Its causes may include individual-level behavioral and psychosocial factors, neighborhood characteristics, environmental exposures…" (Institute of Medicine, 2007).

In an appendix, DJ go on to critique HMS results on prenatal care and smoking:

*D. The results for mechanisms fail placebo tests*

In their Table 3, DJ observe that smoking rates fall for parity 4 vs. parity 3. And DJ Appendix Table A1, they note that when comparing parity 4 vs 3, rates of prenatal care by second trimester or earlier improve, following the 1993 EITC expansion. In both cases they argue that this is evidence against the validity of our research design. We first note that **this is indeed a novel finding** in DJ. Second, we speculate that this finding relates to the observation (documented in HMS, and discussed above) that there was a narrowing of LBW rates comparing parity 4 vs. 3. The timing of the 4 vs. 3 improvements in prenatal care is similar to that of the improvements in low birth weight. We show this below, the red line plots the event study for prenatal care before 3rd trimester and the blue line plots the event study for low birth weight – both comparing parity 4 versus parity 3. We agree that this is a cautionary note – this is why we directly mentioned it in HMS. However, we do not see the addition of smoking or prenatal care as substantively changing the degree of caution required.




**Conclusion regarding mechanisms:**

Clearly, uncovering evidence on mechanisms is important. But, within the HMS paper no one mechanism on its own needs to bear the full weight of explaining the overall impact, and the concerns about a mechanism doesn't detract from the validity or importance of our reduced form results. We speculate that DJ lost sight of why we focused on the mechanisms we examined in the paper: they were the mechanisms available in the birth certificate data. There are other important channels that HMS couldn't address using that data. Two potential mechanisms worth mentioning is that income transfers could reduce stress and improve nutrition. Stress is a documented predictor of pre-term birth (Berkowitz and Papiernik, 1993). There is a growing literature documenting that increases in income can lead to reductions in stress (Aizer et al 2016, Evans and Garthwaite 2014, Fernald and Gunnar 2009 and Haushofer and Shapiro 2016). Of particular note in this context is Evans and Garthwaite (2014) who find that increases in the EITC lead to reductions in maternal stress. Likewise, Lenhart (2019) provides evidence of increases in food expenditures in households exposed to the OBRA93 expansion.

That being said, we agree that a better understanding of the mechanisms behind our reduced form results is an important area for future research.

**<u>DJ Point 4:</u> Difficulty in identifying effect of a national policy at a single point in time (DJ Section IV)**

In HMS we note the importance of potential confounding by other policy or economic variation. As such, we control for several state-by-year varying policy and economics controls. In DJ Table 4, they (1) collapse these variables to create a time series, (2) interact these time-series policy variables with parity indicators, and (3) show that the point estimates and statistical significance of the main EITC effects are somewhat sensitive to inclusion of these policy-time-series-by-parity variables.

First, we note that there is no valid reason to collapse these variables to the annual level. The underlying variation of these policies is at the state-year level. In this sense, **the models in Table 4 are unfounded**. We speculate that DJ may be motivated by one of two factors. (1) they may be concerned that the underlying identification strategy in HMS is a 2xT difference-in-difference strategy; and that strategy relies on a parallel trends assumption. Therefore, they may view inclusion of this new variable (the state-year variables collapsed to a time series) as a test of the parallel trends assumption. The parallel trends assumption is inherent in a difference-in-difference strategy, and we mention this assumption in HMS (HMS p. 174): So, if this is the primary concern, **it is not novel, it was already discussed in HMS**. Separately, we do not believe that the models in Table 4 are a valid test of the parallel trends assumption. Second, we speculate they might be motivated to collapse these variables into a time series because (2) they posit the crack cocaine epidemic is a confounder for the EITC expansion. DJ describe that they only have access to a single time series to proxy for the crack cocaine epidemic. We speculate that only having a single time series may have led to the decision to collapse the policy variables in this case.

Second, we believe the idea of interacting policy variables with parity has merit. HMS only included these state-year variables as additional controls, and did not include their interactions. In a response[5] to an earlier version of DJ, we added interactions between parity and a set of these policy variable controls (as well as other policy variable interactions, see Response Figure 4). The inclusion of these interactions left the estimated treatment effects little changed. This is additional evidence that the conclusions DJ draw from their Table 4 **are unfounded**.

Finally, we note that the discussion in DJ of Table 4 focuses mostly on statistical significance. However, in the horse race specifications, of the 6 coefficients (3 policy interactions x 2 demographic categories): 3 are significant; 4 have meaningfully large point estimates (0.15 or greater in magnitude); and all 6 have confidence intervals that include large effect sizes (0.25 or greater in magnitude). Even taken at face value, this table isn't evidence that "EITC doesn't matter"; it's just mushy and inconclusive evidence. We speculate that this is an illustration of the standard confusion between "absence of evidence" (no stars) and "evidence of absence" (coefficient must be 0). For this separate reason, the conclusions DJ draw from their Table 4 **are unfounded**.

**<u>DJ Point 5:</u> Sensitivity to adding quadratic trends by parity to panel fixed effects model using multiple expansions (DJ Appendix A, referred to in DJ footnote 14):**

While the OBRA93 analysis is at the core of HMS, we also present estimates from a panel fixed effects model to allow for leveraging the variation across multiple EITC expansions. We estimate a parametric model where we control for the maximum credit (varying by parity and year) along with a full set of year and parity fixed effects (as well as demographic cell fixed effects, and state-year policy controls). This model is presented in HMS equation 2.

In DJ Appendix A (pp. 8-9 and Table A2) DJ critique this approach by showing the sensitivity of the results to adding controls for parity x linear trend and parity x quadratic trends.

[5] https://gspp.berkeley.edu/assets/uploads/research/pdf/DJ_response_081018_update_121818_1.pdf

*What is new and what is not new in DJ's critique?* The issue of sensitivity to parity trends is presented and discussed in HMS. In HMS Table 5 we show the sensitivity of the results to adding controls for parity by linear time. In discussing these results, we say:

> "Due to the longer time span, with multiple EITC expansions, we can explore the sensitivity of the results to the inclusion of parity-specific linear trend (in year). The results (in columns 2, 4, and 6) show substantially larger estimates treatment effects for model with parity linear trends. While we may be "overfitting" the parity-time relationship, we view the robustness to including the parity trends as an important result." (HMS p. 195)

DJ expand on this approach and also control for parity by *quadratic* linear time. On page 11 DJ refer to the inclusion of quadratic trends by parity "*given the curvilinear relationship to low birth weight*." Because the majority of the EITC occurs with the OBRA93 expansion, this represents a considerable burden on the data to estimate quadratic trend terms *separately from the effects of the policy*.

Presumably DJ's intention in adding quadratic controls is to control for "pre-existing trends"; trends that would have happened even if the policy had not changed. However, because the policy does change frequently over time, and in broadly non-linear fashion (see Figure 1 above), there is a nontrivial risk that the quadratic trends will be over-fit to capture the impact of the policy. This would not happen if the trends were fit only to time periods with no policy changes. But with a global polynomial fit there is a risk of the trends becoming "bad controls", in the sense of being fit to the causal impacts of the policy changes.

Finally, quadratic trends are not a standard specification in this literature. Given concerns about the short length of the time series, the many periods of change, and the potential for overfitting, it is unclear whether the quadratic should be preferred over linear or no trends.

DJ also present a new analysis limited to data from 1983-1993. They present this as a check against "over-fitting of trends". It is not clear how this analysis speaks to that concern. We believe that it does not address this concern.

Overall, this DJ critique either not new -- **HMS discuss and explore this criticism in their paper** – or it is overfitting the data.

**For readers who are in the weeds on HMS and DJ, there may be things that confuse you in DJ. We enumerate some items that we have noted here:**

- DJ Figure 1 contains a calculation error. The graph includes a label "Δ93-96=$1,283". The label is identifying the change in the maximum credit for parity 3+ between 1993 and 1996 – the correct amount is $1,867 (in 1995 $). The $1,283 is the *difference* in the 1993-1996 change for parity 3 minus the 1993-1996 change for parity 2 ($1867-$583). This number is correctly cited by DJ on p.2: "By 1996 women of parity 3+ were eligible for $1,283 more in tax credits in 1995 dollars than women of parity 2 (Figure 1)."

- DJ middle of page 6: DJ discuss the results for 1987-1990 as the pre period (compared to the pre period in HMS of 1991-1993) – they point out that "The protective effect of the EITC using this pre-period is more than twice as large for all women …" In all results the low birthweight results in HMS we present estimates for a first stage on after tax income to convert the ITT to TOT effects. It doesn't make sense to directly compare ITT estimates (as DJ do here). TOT comparisons is the approach to use.

- On DJ p. 11 refer to the EITC as "*cash transfers from the EITC received mostly in February and spent largely on durables and transportation…*" Patel (2012) finds that the EITC leads to increases in primarily non-durables, work related, and housing expenditures. Increases in durable and investment expenditures are concentrated in the first quarter, around the time of refund receipt. But taking account of the changes throughout the rest of the year, the effects are more concentrated in non-durables.

- DJ sections III (*The Etiology of Low Birth Weight and Plausible Mechanisms*) and IV (*Alternative Explanations*) focuses on "pre-term births" as the object to be explained by the analysis of mechanisms in HMS. However, HMS provide estimates for low birthweight (LBW) status and not preterm birth. While the two are correlated, they are not the same. By moving between discussing HMS LBW results and citing medical literature that references preterm births, the distinction between the two is conflated and it leads to the impression that clinical research that applies to preterm births applies equally to low birth weight.

- If DJ's estimating equations are being attributed to HMS ("HMS aggregate individual-level data … and estimate the following regression (HMS Appendix B)") then the equation should match what is estimated in HMS. They don't because (a) we use Parity2plus not parity 2 as treatment, and (b) DJ include a subscript for demographic group k as a fixed effect when there is no k subscript on the cells defining their dependent variable.

Many typos / errors. A few examples include: (a) DJ top of page 5 – there are two references to "nominal changes" when we think they should say "real changes"; (b) DJ just prior to equation (1), "1981 to 1998" should be "1991 to 1998;" (c) panels are labeled as "Pane" in Table 2; and (d) in DJ Appendix A equation 2 there is an undefined $\varphi_k$, multiplied by the demographic group fixed effect ($\varphi_j$)

<u>References</u>

Aizer, Anna, Stroud, Laura and Stephen Buka (2016) “Maternal Stress and Child Outcomes: Evidence from Siblings” Journal of Human Resources, 51(3): 523-555.

Averett and Wang (2013). “The Effects of Earned Income Tax Credit Payment Expansion on Maternal Smoking,” Health Economics, 22: 1344–1359.

Bitler, Marianne, Hilary Hoynes and Elira Kuka (2017). “Do In-Work Tax Credits Serve as a Safety Net?” *Journal of Human Resources* Vol 36, Issue 2, pp. 358-389.

Cowan Ben, and N. Tefft (2012). “The Effect of Earned Income Tax Credit Expansions on the Smoking Behavior of Women”. The B.E. Journal of Economic Analysis & Policy.

Evans, William N., and Craig L. Garthwaite. 2010. “Giving Mom a Break: The Impact of Higher EITC Payments on Maternal Health.” American Economic Journal: Economic Policy, Vol 6, no. 2: 258–90.

Fernald, Lia and Megan Gunnar 2009. “Effects of a poverty-alleviation intervention on salivary cortisol in very low-income children,” Soc Sci Med. 2009 Jun; 68(12): 2180–2189.

Haushofer, Johannes and Jeremy Shapiro (2016). The Short-Term Impact Of Unconditional Cash Transfers To The Poor: Experimental Evidence From Kenya.

Hoynes, Hilary and Ankur Patel (2018). “Effective Policy for Reducing Inequality? The Earned Income Tax Credit and the Distribution of Income,” Forthcoming, *Journal of Human Resources*.

Hoynes, Hilary and Diane Whitmore Schanzenbach, 2012. “Work Incentives and the Food Stamp Program,” *Journal of Public Economics* 96(1-2): 151-162.

Kearney, Melissa and Phillip Levine (2015). “Investigating recent trends in the U.S. teen birth rate,” *Journal of Health Economics*, 41(2015):15-29.

Kenkel, Schmeiser, and Urban (2014). “Is Smoking Inferior? Evidence from Variation in the Earned Income Tax Credit,” Journal of Human Resources, Volume 49, Number 4, Fall 2014, pp. 1094-1120.

Lenhart, Otto. "The effects of income on health: new evidence from the Earned Income Tax Credit." *Review of Economics of the Household* 17.2 (2019): 377-410.

Meyer, Bruce D., and Dan T. Rosenbaum. 2001. “Welfare, the Earned Income Tax Credit, and the Labor Supply of Single Mothers.” Quarterly Journal of Economics 116, no. 3: 1063–1114.

Nichols, Austin and Jesse Rothstein. 2016. “The Earned Income Tax Credit.” In Economics of Means-Tested Programs in the United States, Volume I, edited by Robert Moffitt. National Bureau of Economic Research Conference Report. Chicago: University of Chicago Press.

Patel, Ankur (2012). “The Earned Income Tax Credit and Expenditures”, U.S. Department of Treasury, November 2012.

Figure 1

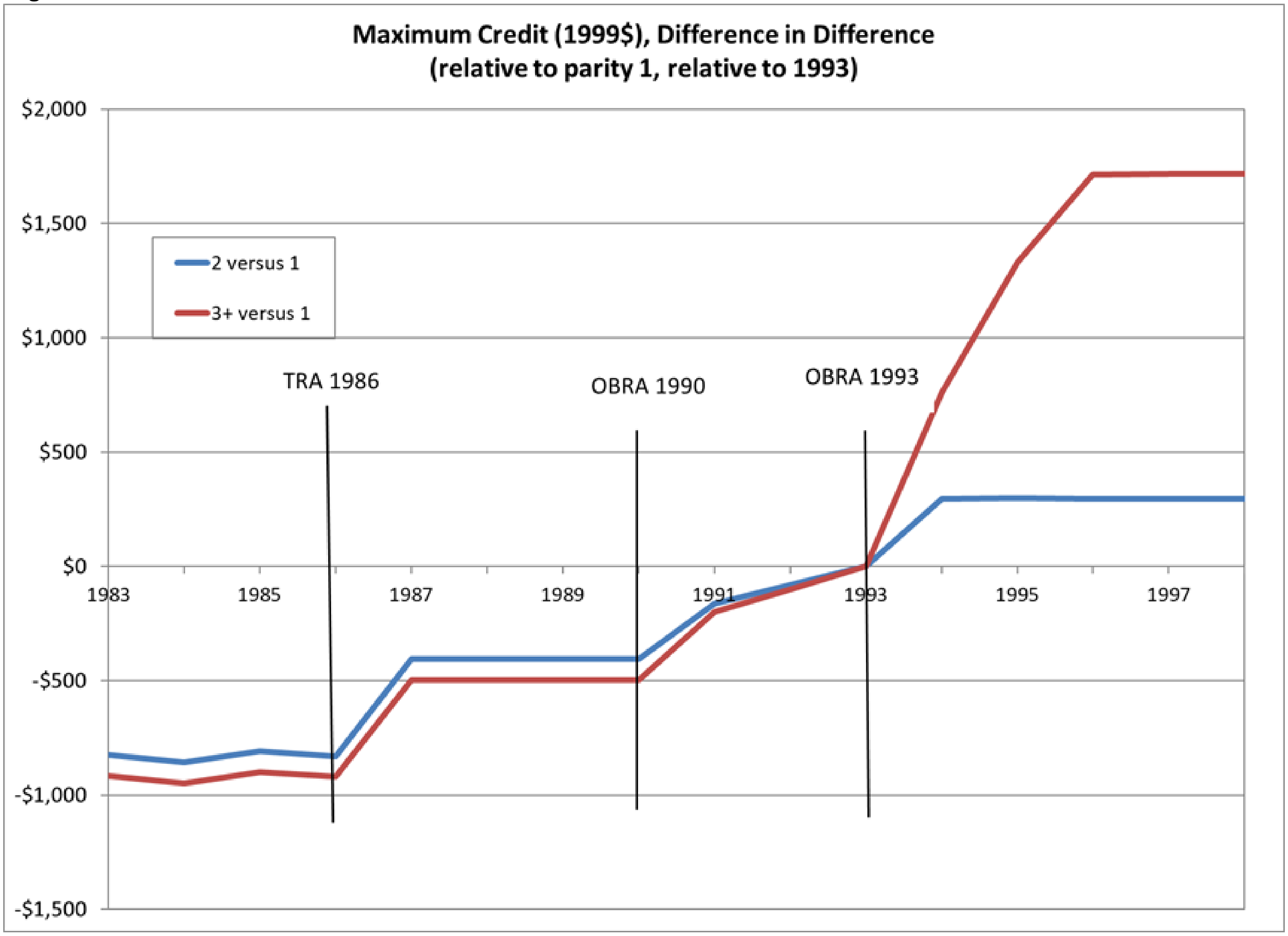

Maximum Credit (1999$), Difference in Difference
(relative to parity 1, relative to 1993)
$2,000
$1,500
$1,000
$500
$0
-$500
-$1,000
-$1,500
1983
1985
1987
1989
1991
1993
1995
1997
2 versus 1
3+ versus 1
TRA 1986
OBRA 1990
OBRA 1993

Table 1: Maximum Credit (1999$) by family size (parity of birth)

| | 1987 | 1988 | 1989 | 1990 | 1991 | 1992 | 1993 | 1994 | 1995 | 1996 | 1997 | 1998 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Level | | | | | | | | | | | | |
| childless (parity 1) | $0 | $0 | $0 | $0 | $0 | $0 | $0 | $347 | $346 | $346 | $347 | $347 |
| 1 child (parity 2) | $1,259 | $1,259 | $1,260 | $1,260 | $1,503 | $1,585 | $1,666 | $2,308 | $2,312 | $2,310 | $2,309 | $2,310 |
| 2+ children (parity 3+) | $1,259 | $1,259 | $1,260 | $1,260 | $1,557 | $1,657 | $1,756 | $2,862 | $3,433 | $3,817 | $3,819 | $3,819 |
| Difference (relative to 1993) | | | | | | | | | | | | |
| childless (parity 1) | $0 | $0 | $0 | $0 | $0 | $0 | $0 | $347 | $346 | $346 | $347 | $347 |
| 1 child (parity 2) | -$407 | -$407 | -$406 | -$406 | -$163 | -$81 | $0 | $642 | $646 | $644 | $643 | $644 |
| 2+ children (parity 3+) | -$497 | -$497 | -$496 | -$496 | -$199 | -$99 | $0 | $1,106 | $1,677 | $2,061 | $2,063 | $2,063 |
| Difference in difference (across parity, relative to 1993) | | | | | | | | | | | | |
| 2 vs 1 | -$407 | -$407 | -$406 | -$406 | -$163 | -$81 | $0 | $296 | $300 | $297 | $296 | $297 |
| 3+ vs 1 | -$497 | -$497 | -$496 | -$496 | -$199 | -$99 | $0 | $760 | $1,331 | $1,714 | $1,717 | $1,716 |
| 3+ vs 2 | -$90 | -$90 | -$90 | -$90 | -$36 | -$18 | $0 | $464 | $1,031 | $1,417 | $1,421 | $1,420 |