

Replication Section

Response to Breusch and Gray

Bruce Chapman and Matthew Gray

1. Background and introduction

In July 2004 the world record holder and Sydney Olympic gold medallist for the men's swimming 400 metres freestyle, Ian Thorpe, competed in the Australian trials for the Athens Olympics. Unfortunately, he slipped off the starting blocks, this apparently meaning that he became ineligible to compete in the event at the Olympics. The incident, and its apparent consequences for Thorpe, his country, and the sporting public generally, were portrayed in Australia as a minor disaster for humankind, and a major tragedy for domestic sport.

For many days in Australia after this incident the media lamented the rules, crying that it was unfair that "our Thorpie" would not be able to compete in his best Olympic event. The relevance for the current exercise is that Ian Thorpe later recounted that on the night after this traumatic incident he received an email from a very close friend. It said, simply, "whoops".

Not long before this Trevor Breusch came to us and explained that he and Edith Gray had spent a lot of time examining our paper (Chapman et al. [4] (hereafter referred to as CDGLM)) and had found some points of confusion. We were asked if they could have access to our data and the computer programs used in the generation of our results. We readily agreed and supplied all that was asked; meanwhile we felt a bit bothered by the possibility that the paper might have mistakes.

After a short time Trevor came to visit again and revealed (politely) that, yes indeed, our paper did have mistakes. Some of the errors seemed to be the sorts of things that are not uncommonly present in applied empirical economics, such as with respect to the programming of the data. But at least one mistake seemed to be quite clumsy, even embarrassing.

After we had the chance to digest the gist of their critique our reaction was the same as Ian Thorpe's friend: "whoops". It was apparent that we had metaphorically fallen off the blocks. Unlike Ian Thorpe we don't have a world swimming record to defend, but we care very much about our reputations as social scientists.

This journal is motivated by the importance of honesty and co-operation with respect to the replication of applied empirical analysis. This is a perspective that we

endorse completely, and it is one that underlay our willingness to provide all of the technical data and programming material relevant to the paper scrutinised by Breusch and Gray. But we would be less than honest if we did not admit to feeling, initially, that we would have preferred for this not to have happened.

Trevor Breusch told us a few months ago that the Breusch and Gray critique of our paper would likely be appearing in this journal, and that the editor would welcome a response from us, which now follows. It takes the following forms.

First, we accept in general the accuracy of the Breusch and Gray documentation of the errors in our paper, and have chosen not to go into the myriad details of their critique of our work. Second, we take the opportunity to examine the underlying meaning of estimates of the forgone earnings from child rearing. Third, and most importantly, we examine empirically several aspects of the methods typically employed in the estimation of the forgone earnings from child rearing.

The important bottom line that became apparent through this process is that we now believe that the methods and data typically used in this area seem to have fundamental conceptual and empirical weaknesses.¹ We have become convinced that even with the technically careful recent contribution of Breusch and Gray [2] (which won a best paper prize), at this stage applied economics has not delivered reasonable estimates of the forgone earnings from child-rearing, and it might even be the case that it is unable to do so with current approaches and data.

All social science research is a process, not an ending, and this important lesson for us has been clarified and reinforced very productively in this (somewhat uncomfortable) journey. A critical point is that, in the absence of the Breusch and Gray criticism of our work, and its encouragement and endorsement by economic journals such as the *Journal of Economic and Social Measurement*, we would not have walked down this road.

2. What is the economic meaning of the forgone earnings from child rearing?

The Breusch and Gray critique of our work (hereafter referred to as BG) encouraged us to reflect on the meaning and importance of estimates of the forgone earnings from child rearing. This might seem to be a strange thing to now contemplate for at least one of us (Bruce Chapman), since he first researched the area as long ago as 1988. But the question arose with the observation from BG that “The findings of this research [CDGLM] have been given extensive coverage in the popular media in Australia, both at the time of its publication and afterwards whenever female participation in the labour force becomes a matter for policy debate or general public discussion” (BG, p. xx).

¹Our comments apply to much of the work we know in this area: Beggs and Chapman [1]; CDLM and Breusch and Gray [2].

BG implied that our results were accepted publicly as a true statement of the forgone earnings from child rearing, and that the findings had some significance for policy debate. While it is certainly true that our results were reported in the Australian media at the time, it is also very unlikely that our analysis had any consequences for policy debate. The reason is that such estimates reflect one aspect only of family economic circumstances, a point we made in CDGLM.

On the above we stated: “The broad point is that estimates of the ‘forgone earnings from child rearing’ should not be interpreted only as the market opportunity cost of children for the general population. Fertility is not an exogenous event, and work decisions reflect highly significant factors that are very difficult to model and control for. Further, whether or not to have children, and the number to have, depends on other household choices, and these are not modelled.” (CDGLM, p. 376).

There are two points concerning the lack of direct policy relevance of estimates of forgone earnings. The first is that decisions within a household concerning child-rearing arrangements will be made in the context of the economic behaviour of all members of a household. Specifically, if a mother chooses to not participate in the paid labour market for a period following the birth of a child, such a decision necessarily impacts importantly on other economic decisions of that household. As examples: the hours and effort supplied in the labour market by a father; household savings behaviour; and consumption.

Second, levels of forgone earnings are not meaningful as measures of, or proxies for, welfare. Mothers and families might have high or low forgone earnings from child rearing, as a consequence of shadow wages and/or labour supply decisions, but a given level of forgone earnings will not obviously be related to poverty incidence, deprivation or wealth. And for policy there are no obvious distributional implications, as there would be for example with levels of household income.

Thus, with respect to public policy debate, estimates of the forgone earnings from child rearing considered in isolation from other household circumstances are not very interesting. This raises the broader question: with respect to family policy issues, why have economists (including us) been involved in empirical exercises of this form? We do not have a compelling answer to this line of enquiry, and our reservations concerning the relevance of the issue are reinforced below with respect to the empirical problems associated with the empirical methods employed in the area.

An important point from the above is that, before BG, we had not been sufficiently critical of the underlying economic basis of estimates of the forgone earnings from child rearing. This was an important first lesson from the reflections demanded of us. There are several more.

3. The fragility of forgone earnings estimates: Functional form

A main conclusion from BG is that the CDGLM estimates of forgone earnings are in error. As we have confessed above, we now know with certainty that programming

and other mistakes make this true for the dollar values of forgone earnings, although not significantly so for the proportions of a childless woman's earnings associated with child rearing. But in the process of revisiting our estimates, and those of others, it became apparent that in this area (and possibly most other applied economics topics) there are myriad ways of producing widely misleading estimates of relationships beyond the fairly creative errors now identified in CDGLM. And it is likely that many of these ways exist in all current papers in the area.

To illustrate the potential to make invalid inferences in ways unrelated to programming and proof reading, let's take the issue of functional form. A key issue in all empirical work is the choice of functional form, more specifically, the restrictions placed upon the estimated coefficients.

In the Australian forgone earnings research, such as reported in Beggs and Chapman [1], CDGLM and Breusch and Gray [2], female labour supply and wage models are used to determine the effect of the presence of children on a mother's total earnings. In all of this research it is found that a woman's level of education is a very important explainer. Further, it has always been found in proportionate terms that the higher is the level of education the lower are the forgone earnings associated with children. Indeed, the result has become a stylised fact of this research, leading Breusch and Gray [2, p. 140] to state: "A common finding in all the previous studies is that women with higher levels of education lose less of their earnings *in proportional terms* due to motherhood than women at lower levels of education. That finding is repeated here".

However, all the research imposes strong restrictions on the role of education in these relationships. In particular, the estimation approaches constrain the coefficients of all other variables to be identical irrespective of the level of a woman's education; this means, for example, that estimated changes in earnings with additional years of age and labour market experience would be held to be the same between women with university degrees and women who had not completed high school. In retrospect this seemed to us to be very unlikely to be true, and we wondered if the stylised fact concerning the role of education would be sustained through a relaxation of the restrictions.

In order to illustrate the sensitivity of the estimated foregone earnings to the education restriction we have reworked the estimates of Breusch and Gray [2] using the Household, Income and Labour Dynamics Australia survey (HILDA), 2001.² We have chosen to explore these relationships with extensions of their model not because they are our critics, but because this is the most recent work in the area and because HILDA has excellent features for applied micro-econometric research. We emphasise that all previous Australian research imposes the same restrictions on the role of education. Our point is generic.

²We are grateful to Trevor Breusch and Edith Gray for supplying all the code and spreadsheets for this and other exercises reported below.

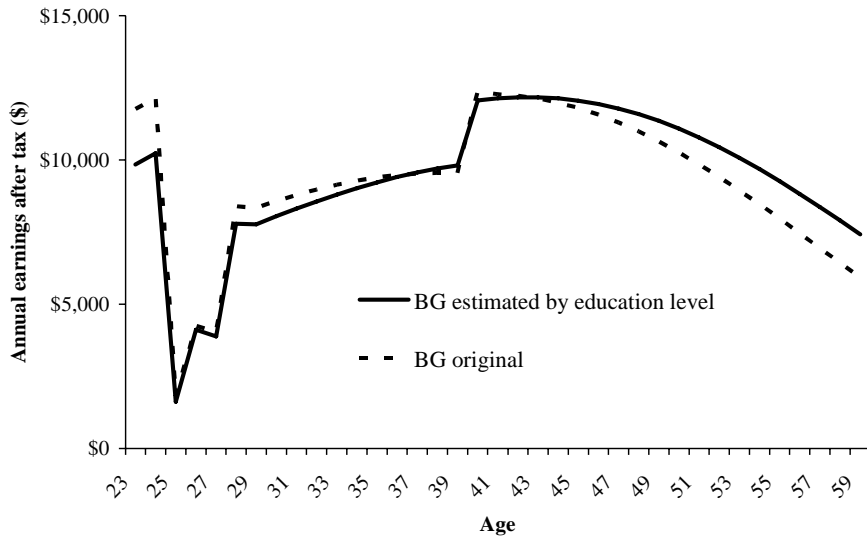


Fig. 1. Lifetime earnings profile of mothers with one child, incomplete high school.

Our approach is to estimate the Breusch and Gray model separately by education level (samples with incomplete high school, year 12 education only, a trade qualification and a degree level qualification), allowing the effects of the non-education explanatory variables to differ between education levels. Forgone earnings estimates are likely to be sensitive to the restrictions that we have relaxed for several reasons. One is the strong possibility that the returns to labour market experience differ according to education. This could be the result of more educated women choosing careers with relatively high levels of on-the-job training, meaning that for given measured levels of experience the actual quality of the accumulated skills of more educated women is higher.

Figures 1 to 3 compare the simulated earnings profiles resulting from the pooled sample (labelled BG original) and those resulting when estimating the models separately by education level (labelled BG estimated by education level). It appears that the relaxing of restrictions on the coefficients of all explanatory variables for each level of education resulted in reasonably close similarities in simulated lifetime earnings profiles to those estimated using the sample pooled across educational attainment.³ However, it turns out that in critical respects there are some important differences. For simplicity the simulations are shown for women with one child only, and for three education levels (incomplete high school, year 12 only and having a degree).

³For reasons of space we show only the simulations. The regression results, data and programming are available from the editor and the authors.

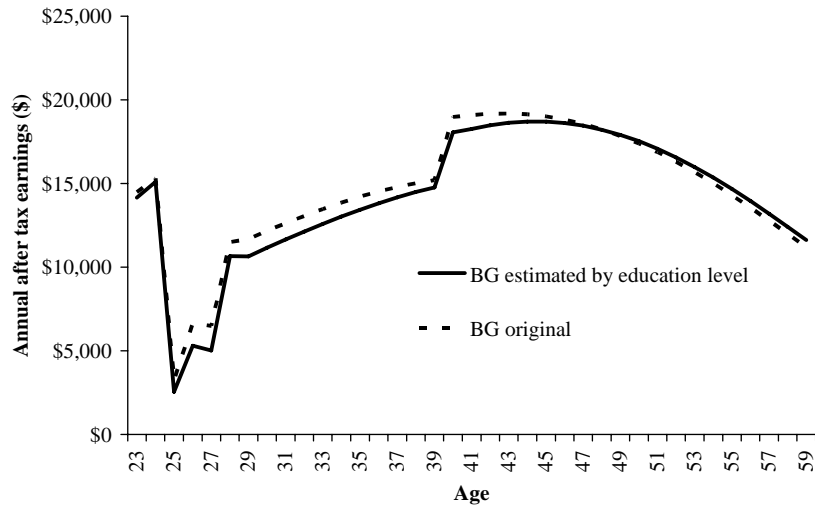


Fig. 2. Lifetime earnings profile of mothers with one child, year 12 education.

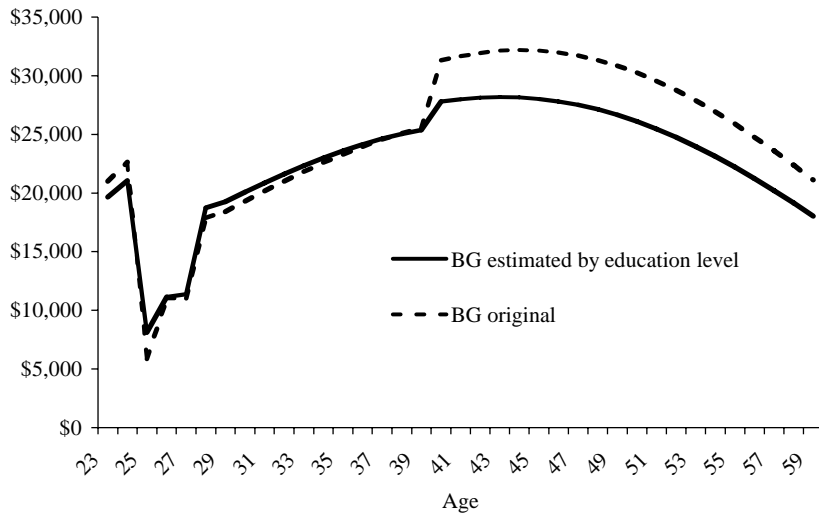


Fig. 3. Lifetime earnings profile of mothers with one child, degree.

The data can be used to show the estimated differences in lifetime earnings, as a proportion of the earnings of childless women, for both the original pooled method and for our less constrained approach. The results are shown in Table 1, with some highly interesting findings.

One is that for women with year 12 and university degrees, the rankings are the

Table 1
Proportionate lifetime earnings of women, pooled and unrestricted comparisons, by education level and number of children

Scenario	Breusch and Gray Original	Breusch and Gray Unrestricted
	Per cent of earnings compared to childless women	Per cent of earnings compared to childless women
<i>Degree</i>		
1 child	72	77
2 children	62	68
3 children	55	63
<i>Completed Year 12</i>		
1 child	69	63
2 children	56	51
3 children	47	42
<i>Incomplete High School</i>		
1 child	60	72
2 children	42	46
3 children	32	34

same across the different methods but the orders of magnitude are dissimilar, and the changes are in diverse directions. For the former education group, there are about five percentage point differences (about 10 per cent higher) in forgone earnings for each of one, two and three children using our unrestricted method. But for the most highly educated women, the changes are in the opposite direction: the unrestricted method provides forgone earnings estimates that are systematically lower.

These results must affect the estimated differences of forgone earnings comparisons made *between* educated groups. For example, in the case of one child, the forgone earnings difference between degree holding women and year 12 women is 3 percentage points only (72–69) with the original pooled method, but the figure increases very substantially to 14 percentage points (77–63) with our unrestricted approach.

However, the most important change arises with respect to women with the lowest level of education. The pooled sample suggests a very much lower estimate of forgone earnings for women with one child, decreasing from 40 per cent (100–60) to 28 per cent (100–72) in the move from the pooled to the unrestricted estimation.

But the biggest difference arises in the forgone earnings comparisons between those with year 12 and those with incomplete high school in the case of 1 child. The unrestricted method now reveals a *fall* in forgone earnings as education level decreases, with the pooled approach providing a difference in the estimates of plus 9 percentage points (69–60), compared to a difference of minus 9 percentage points (63–72) in our unpooled specification. Thus the move to a more flexible and preferred approach overturns the alleged stylised fact that forgone earnings from child rearing always fall in percentage terms as education increases.

The importance of this finding is that important differences can arise in the estimation of the relative size of relationships even when technical approaches are sound.

There is a clear need for all work in this area, including our own, to be more circumspect about method and to put more emphasis on the sensitivity testing of the role of functional form.

4. Empirical and conceptual issues: The counter-factual

The essence of the method used to estimate the forgone earnings from child rearing is as follows. The earnings of mothers are compared with the earnings of otherwise similar childless women. Empirically this is achieved by estimating an earnings function and a probability of employment equation augmented with variables capturing the number and age of children, taking account also of the earnings effect of other measurable characteristics.

A possible weakness in the approach is that when it comes to estimating economic and social relationships for individuals there will be a whole host of unmeasured characteristics that matter for earnings. Some of these will be associated with measured variables influencing labour market and fertility choices. If these unobserved characteristics are correlated with variables included in the analysis it is likely that these have the potential to invalidate the interpretation of measured earnings differences as reflections only of the presence and number of children.

An example of such a factor is that the decision to have one or more children may reflect a woman's interest in and capacity to spend time childrearing versus time spent in the labour market. In other words fertility decisions reflect women's innate abilities and interests. From this perspective, women best suited to labour market activity will take this into account in choices made with respect to children. But there are many other possibilities, and some of these might mean that the biases in the estimations go the other way.

The essential point is that there are likely to be differences between women who choose not to have children, and mothers. And if there are systematic differences between members of the two groups of women it is highly plausible that as well as in fertility decisions they will manifest themselves in labour supply decisions and work effort behaviour.

In order to illustrate the possible importance of the above possibility, we have estimated the Breusch and Gray [2] model separately for the sample of women who are childless and the sample of women who are mothers.⁴ The reasoning behind this method is that if there are unmeasured associations between various personal characteristics and the included variables, the approach adopted would be expected

⁴Apart from allowing all coefficients to vary, the only difference in the specification is that for the childless women the variables relating to children are excluded and for the women with children the variable measuring ever having had a child is excluded since there is no variation for these variables within the respective sub-samples.

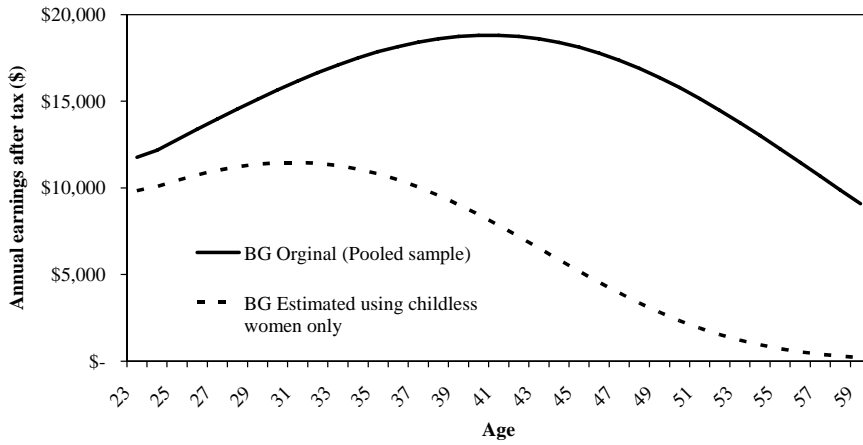


Fig. 4. Lifetime earnings profile of childless women with incomplete high school.

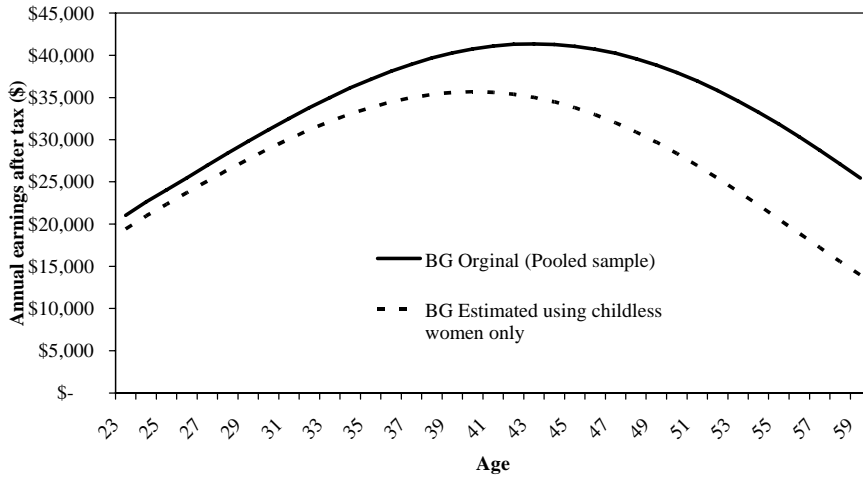


Fig. 5. Lifetime earnings profile of childless women with a degree.

to indicate this indirectly since we are now allowing the effects on earnings of non-children variables to vary between childless women and others. Thankfully with this approach the sample sizes remain substantial, there being 1,489 childless women and 3,229 mothers.

To show how much the constrained regression approaches matter for lifetime earnings, Figs 4 and 5 show the simulated lifetime earnings profiles for childless

Table 2
 Mother's earnings estimates as a proportion of childless women's
 earnings, by education level (one child only)

	Breusch and Gray original %	Breusch and Gray modified %
Degree	72	83
Year 12	69	83
Incomplete high school	60	124

women (with either incomplete high school or a university degree)⁵ for the two different approaches. There are several important points.

The first is that for both women with incomplete secondary and women with a degree the earnings profiles based on the pooled sample and the sample estimated separately for childless women are extremely different. For both education levels the profile constraining relationships to be the same for childless women and mothers are very much higher than the unrestricted estimates. Further, these differences increase with age, suggesting that the women who are childless over their lifetimes have substantially lower earnings than those younger women identified as being childless but who might choose to be mothers in the future.

Since the estimates of lifetime earnings are so different when the sample of childless women is treated separately, it follows that there is a very significant potential for this to impact on our calculations of forgone earnings from child rearing. This is confirmed with the data of Table 2, which shows the earnings of mothers as a proportion of childless women's earnings (for simplicity, for one child only), by education, using the two different approaches. The change in the story is extraordinary.

In particular, treating childless women empirically as being quite different to mothers reveals that the estimates of the forgone earnings from child rearing become substantially lower, by at least 35 per cent, for those with degrees and those with year 12 education only. But the change for the women with the lowest levels of education is the most dramatic and important: estimating the relationships to reflect the possibility that childless women are quite distinctive *reverses* the estimate of the forgone earnings from child rearing.

That is, having one child compared to remaining childless *increases* the lifetime earnings of women with low levels of education, by about 25 per cent. A plausible interpretation of this finding is that there are important unobserved characteristics operating behind the scenes, and these differ very significantly between childless women and mothers (for those with low levels of education). If this is the case it must mean that members of the former group have very poor labour market outcomes for reasons not explained in the data. Of course, the new results might also reflect weaknesses in the method associated with separate estimations for childless

⁵The estimation results are available from the editor and the authors.

women, arising perhaps from sample size issues; but it is difficult to believe that the unobservable variables are unimportant.

The critical bottom line is that the method used to estimate forgone earnings from child rearing, comparing the lifetime earnings of childless women and mothers, apparently has significant empirical weaknesses. A plausible conclusion is that the data are not rich enough, and the methods used thus far have been inadequate.

That is, for members of at least some groups the potency of unobserved characteristics might well be sufficiently high to compromise importantly the efficacy of the exercise since the counter-factual is just not being described sufficiently properly. This is another way of suggesting that fertility choice is endogenous, or at least powerfully reflects selection processes.

5. Conclusions

Although this has been an uncomfortable experience for the authors, it was eventually a very positive one. The importance of replication in uncovering errors in research has been illustrated consummately by BG. We are grateful to Trevor Breusch and Edith Gray for the professional way in which they handled the difficulties in our original paper.

Having been compelled to address the errors identified by BG has meant revisiting some important issues in the area of estimating the forgone earnings of childrearing. We have explored the sensitivity of forgone earnings estimates to the restrictions on the role of education in the relationship between childbearing, employment and earnings made in the existing literature and have demonstrated that some of the key results are quite sensitive to imposed coefficient restrictions. This finding illustrates the need for all work in this area, including our own, to be more circumspect about method and to put more emphasis on the sensitivity testing of the role of functional form.

We have also explored empirically the validity of using the earnings profiles of childless women to obtain the counterfactual estimate of what mothers would have earned had they never had children. One conclusion from our re-estimations is that the method used to estimate forgone earnings from child rearing, comparing the lifetime earnings of childless women and mothers, has apparently important methodological weaknesses. There is a case to be made that the data are not rich enough, and the methods used thus far are not yet adequate. There remain other possible interpretations.

This journal plays an important role in encouraging honesty and co-operation with respect to the replication of applied empirical analysis. We endorse the aims of the *Journal of Economic and Social Measurement* and its replication section editor, B.D. McCullough.

References

- [1] J. Beggs and B. Chapman, *The forgone earnings from child-rearing in Australia*, Discussion Paper No. 190, Centre for Economic Policy Research, The Australian National University, Canberra, 1988.
- [2] T. Breusch and E. Gray, New estimates of mothers' forgone earnings using HILDA data, *Australian Journal of Labour Economics* 7(2), 125–150.
- [3] T. Breusch and E. Gray, Replication: a study of mothers' forgone earnings in Australia, *Journal of Economic and Social Measurement* (2005).
- [4] B. Chapman, Y. Dunlop, M. Gray, A. Liu and D. Mitchell, The impact of children on the lifetime earnings of Australian women: Evidence from the 1990s, *Australian Economic Review* 34 (2001), 373–389.