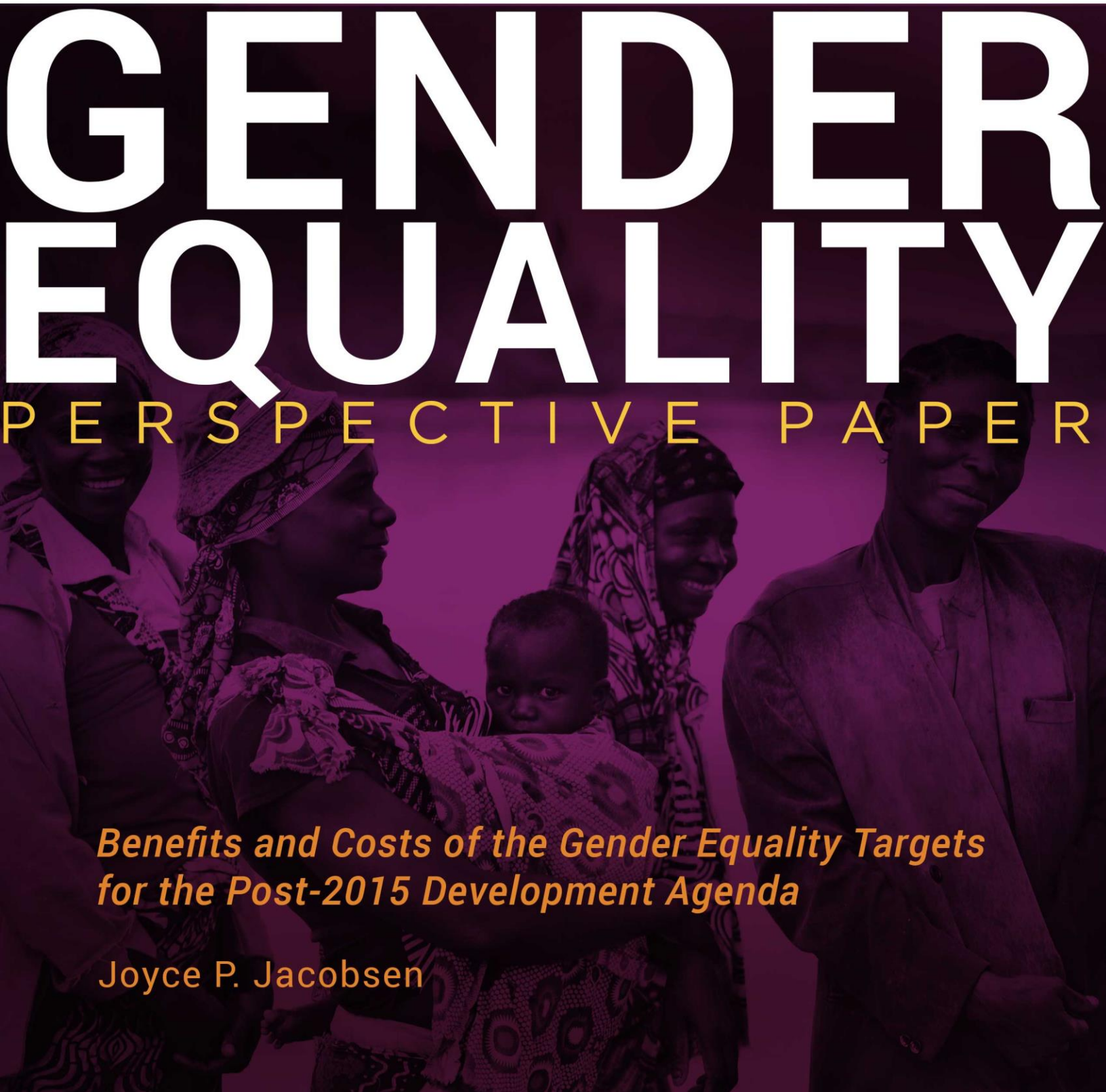
COPENHAGEN CONSENSUS CENTER

GENDER EQUALITY

PERSPECTIVE PAPER



Benefits and Costs of the Gender Equality Targets for the Post-2015 Development Agenda

Joyce P. Jacobsen

POST-2015 CONSENSUS

Benefits and Costs of the Gender Equality Targets for the Post-2015

Post-2015 Consensus

Joyce P. Jacobsen
Wesleyan University

Working Paper as of 9 October, 2014

INTRODUCTION **1**
THE ADVANTAGES AND LIMITATIONS OF BCR ANALYSIS BASED ON SOCIAL EXPERIMENTS **1**
ALTERNATIVE APPROACHES TO SOCIAL EXPERIMENTS FOR GENERATING BCRS **4**
THE ATTRIBUTION PROBLEM **6**
THE LIMITS OF POSITIVE ANALYSIS **7**
CONCLUSION **8**
REFERENCES **10**

Introduction

I enjoyed reading the gender equality assessment paper by Irma Clots-Figueras and found it to be very informative. Clots-Figueras addresses all of the relevant Post-2015 goals (i.e., those having anything to do with gender equality) and reports on the most recent relevant experimental results having to do with explicit benefit-cost ratios (BCRs) for this gender-related research. The paper does not overreach in reporting on these results, and is quite honest regarding cases where the BCRs are low as well as regarding goals where there is little relevant research of this experimental nature.

However, the Clots-Figueras paper essentially limits possible evidentiary sources to experimental papers, and mainly to such studies that explicitly address BCR measurements, rather than attempting to extrapolate from nonexperimental papers and other forms of research to see what more can be learned. This is not so much a weakness of the Clots-Figueras paper so much as it is a purposefully-made choice of how to address the topic at hand, which serves in part to keep the project to a manageable length. Thus that choice also leaves me plenty of ground to comment in this response on these alternative complementary approaches to assessing the effects of policy changes on measures of gender equality.

Hence my perspective paper will not highlight additional targets or issues of the challenge that were not sufficiently dealt with in the Clots-Figueras paper (as the paper did appropriately lay out the relevant challenges). Instead, it will suggest different views regarding how to measure and assess evidence regarding BCRs. This includes addressing the question of what elements are preconditions and what events are actually more likely to occur if economic situations for women are improved.

The Advantages and Limitations of BCR Analysis Based on Social Experiments

The Clots-Figueras paper ably demonstrates both the advantages and limitations of BCR analysis as currently practiced when applied to gender issues. These include a reasonable way of comparing different policies by comparing their BCRs, but also concerns regarding the interpretation of BCRs for policy setting, and concerns regarding the broader applicability of project-specific BCRs scaled up into a larger context.

As with all BCR analyses, some questions come to mind regarding how to view these BCRs: in each particular case, should the BCR be viewed as a lower bound, an upper bound, or as a point estimate with a confidence interval (of some unspecified size) around it? Also, in cases where BCRs are low, is this because benefits are not sufficiently accounted for (and where BCRs are high, is this because costs are not sufficiently accounted for)? Particularly in cases where there are many effects that may accrue to persons other than the treatment candidate, the benefits and costs may be hard to measure. The spillover effects of changing one woman's life by, say, increasing her education or health care, on family, coworkers, and neighbors may be substantial.

If we set the high standard that the effects solely on the individual person affected directly by the policy have to meet a high BCR standard, we may be foregoing activities that actually have a very high societal BCR. Alternatively, are we ever going to be able to rule anything out definitively if it appears to have a low BCR, given that there may be these large unmeasured spillover effects?

BCR analysis traditionally works best and is most convincing in cases where it is straightforward to identify the beneficiaries and the effects of the intervention, and the main question is which way is it more cost-efficient to deliver the effects. For example, with medical analyses of early intervention vs. later treatment for a given disease, it is easy to identify the costs of treating the same disease in two different ways, with payoff in both cases being the avoidance of disease and the reduction of the years of suffering for the individual. So BCR analysis appears to be most convincing and useful when multiple methods of achieving the same desired effect are contrasted.

However, for gender-related outcomes, it is both unclear what the range of effects may be attributable to a given policy intervention, and whether all of those effects are desirable. Contrast the preceding medical example with an example involving gender empowerment such as increasing education, where multiple additional outcomes are set into play. There are effects on both market and nonmarket activities, on fertility, on lifespan, and on the wellbeing of other members of the household. And some aspects of gender empowerment do not even involve such a clear-cut gain for the individual before moving to consideration of additional outcomes. For example, increasing women's labor force participation is not necessarily an end in itself, and also such an increase may be accompanied by other outcomes that may be viewed as neutral, desirable or undesirable. Increases in women's labor force participation correlate with a number of other societal outcomes, including decreased rates of marriage, reduced fertility rates, and increased demand for child care services. These other changes may cause additional requests for social services (such as increased demand for state-subsidized child care) that lead to additional expenditures by the government. And reduced fertility rates, while potentially viewed as socially desirable if they drop down to where the society is at replacement level, may be considered problematic if they drop further to below population replacement level. Such a drop and the associated aging of the population can again lead to additional calls for state-supplied services such as elder care.

An additional set of concerns arises in considering how project-specific BCRs scale up to society-wide policies. Here the standard concern of economists regarding partial vs. general equilibrium analysis arises; true, this is a concern that most economists state in passing before indicating that they will nonetheless be using partial equilibrium analysis, but here it really does appear to be relevant.

A specific example, of great interest to many relatively developed countries: increased higher education sounds like a good thing—so long as the jobs are there for these higher educated persons. Otherwise concerns about overeducation start to arise. Skill mismatch problems can be quite severe. Thus, while educating one person all the way through college can have a significant payoff for that person, educating a large number of people

may reduce their collective outcomes significantly relative to the outcome for the one person, both in terms of employment probabilities and salaries. Another example: using microfinance to improve one person's business prospects in an emerging area (such as providing satellite phone service to villagers), vs. adding another entrant into an already-crowded market (such as selling produce in a current marketplace with plenty of existing vendors). It is important also to measure crowding-out effects on other groups besides the directly affected group. For example, if women's employment rises, does men's fall as a direct result of that rise? While this still can mean that society as a whole is better off (men may choose to reduce employment due to the income effect through the increased earnings of women in their households), not adjusting for the offset means an overestimate regarding the policy's effect on total earnings for households.

With small-scale experiments it is often also unclear practically how they would scale up. For instance, with the example given in the paper from Uganda of after-school clubs for girls that provide "both vocational training and information on marriage, sex and reproduction," would such clubs be made available (or even mandatory) in all Ugandan communities, both rural and urban? And would such clubs still be considered useful and distinctive if all communities had them (as opposed to a "specialness" effect from being the community that has such a club as opposed to, perhaps, the nearby community without one).

While the preceding examples illustrate how partial equilibrium analysis can overstate general equilibrium returns, the opposite case can also hold: partial equilibrium analysis may understate returns if a society-wide innovation sets off a series of systemic changes with much greater effect. For instance, one well-educated person may not be able to have much influence on existing social structures (even as they may still do well personally in those structures), but a more educated population may lead to more efficient institutions and more efficient interactions with others, thus leading to lower costs for the society as a whole for doing business.

The other main concern of using social experiments as the main way of measuring gender-related changes is that many topics cannot be easily addressed by this format. Any type of larger social change in particular is hard to simulate using this methodology, as are changes that involve meddling into individual lives at a level that is considered socially inappropriate or unethical. Yet gender is so basic an element of human life, as well as so pervasive an element, that any targeting of gender-specific society-wide outcomes almost unavoidably involves large-scale social change.

Thus the Clots-Figueras paper underscores the fundamental limitations of the small-scale experimental approach and the potential pitfalls of overreliance on these studies alone in trying to determine the most likely effects of changes meant to affect gender dynamics in societies.

Alternative Approaches to Social Experiments for Generating BCRs

Given that I have now laid out a case for why BCRs meant to be used as policy selection guides cannot be based on social experiments alone, what other approaches can complement social experiments and lead to alternative calculations of BCRs? These alternative BCRs then can be viewed both as robustness checks on those generated from social experiments, as well as potentially preferable calculations if their calculation methodology is viewed as more convincing, perhaps because they can address more convincingly one or more of the problems outlined above.

BCRs can be generated from any empirical case where one can contrast two or more situations that allow for ceteris paribus comparison by holding at least some factors constant (or in common), and where there is some knowledge of the relative costs of entering the two situations.

For instance, take a set of countries (or a pair of countries) that have different levels of social spending on a particular matter of interest, such as women's health care. We can then compare the different outcomes for such standard indicators as maternal mortality, infant mortality, and other health outcomes for both women and others (especially children). Combined with calculations of the monetary value of a DALY, this can then lead to a BCR calculation.

This is really a very standard approach for seeing what differences in treatment lead to differences in outcome. It is generally the case that the study does not go the next step towards generating a BCR, but the evidence from the study can still be used ex post to generate a BCR. And often there is even a regression coefficient that gives a direct measure of the marginal cost of generating another unit of an outcome. This marginal calculation may be preferable to the averaged calculation otherwise generated, as it shows the BCR for the next unit of input (or dollar) provided.

Where can these situations arise? They can include: 1) current or historical examples taken as natural experiments (where two or more societies are compared, with the view that some important aspects of their situation are held in common or held constant; 2) current or recent examples presented in a cross-sectional or panel multiple regression framework (where sufficient data are collected for the societies in the sample to allow for various factors to be treated as if held constant); and 3) historical examples involving evolution along a time line (so that the same society is compared at different points in time, again assuming that this allows critical aspects of the situation to be held constant over time)

Some of the most interesting historical examples involve two neighboring countries, often with a common cultural heritage, that then separate in terms of their social policy at some critical juncture. One interesting such natural experiment was the division of Germany into East and West following World War II up until the early 1990s, where one side was subject to standard capitalist development while the other side followed socialist doctrine. In this

situation, the gender outcomes ended up being quite different between the two sides, including much higher women's employment rates on the Eastern socialist side, as well as higher rates of child care provision and lower fertility rates. The second interesting aspect of this natural experiment was the reunification of Germany and the subsequent reequilibration of the two sides, whereby the higher rates of female participation then fell sharply. In both cases, these changes could be used to generate BCRs given some rudimentary data on differences in social spending across jurisdictions.

Countries in disparate locations but with a common cultural background, such as the former UK colonies, also make for a useful natural experimental setting—indeed, a popular comparison set is Australia, the US, and the UK itself. One interesting study showed how differences in timing and implementation of equal pay measures had notable effects on the gender wage ratio viewed in time series across the three countries (Gregory et al., 1989). Differences in spending on implementation (including willingness to spend on enforcement) could thus be compared to the benefits accruing to women of higher pay to generate comparable BCRs for the period across countries.

Many policies can be studied in a BCR framework using cross-country (and increasingly panel) data that are collected by the international agencies such as the World Bank (2014), IMF (2014, and the United Nations (2014). All of these agencies are essentially creating multivariable, multicountry, multiyear databases that are increasingly conducive to multiple regression analysis. In particular, differences in social spending, including spending on different specific areas, can be used as an independent variable and a variety of dependent variables can then be regressed on spending and control variables to calculate the marginal effects of changes in spending.

In addition, the increasing availability of household expenditure surveys for many countries allows a similar BCR analysis to be done within country using a household as the unit of observation. Here, similar to the above society-wide studies, spending on different household expenditures can be used as independent variables and various household outcomes can then be regressed on expenditures to calculate the marginal effects of differences across households in spending.

Turning to time-series historical examples, one of the most interesting case studies is Sweden, where the various changes in social and family policy, including tax code changes, have been clearly documented by Statistics Sweden and other sources. The post-World War II period has also seen enormous changes in Swedish family life, including a large rise in cohabitation relative to marriage, and changes in fertility rates, where the time series of these changes can be linked to specific policy measures (which again have specific costs associated with them, such as the very large costs of increases in paid parental leave).

In a larger sense, the whole path of human development is evidence in favor of the approach of increasing women's equality as it is clear from cross-country panel data and from within-country observations over even longer time spans that economic development and women's equality almost always occur in near tandem, with more developed countries also having higher levels of gender equality as measured by a variety of indexes created by

various organizations (a notable exception perhaps being Saudi Arabia) (World Economic Forum 2013). So here the benefit is measured by higher levels of GDP per capita, while the costs are such measures as the higher cost of educating women and expanding the legal system to monitor and adjudicate violations of their rights. Even if the causality is not clear, why not take a chance on higher development by expanding women's rights, capabilities, and resources?

In addition to these conventional approaches as outlined above, there are additional more novel approaches that can be used to generate BCRs, again using essentially a natural experiment framework. For one thing, we observe migration patterns by gender, both within and between countries, as examples of people voting with their feet as to where they would prefer to live and in general carry out economic activities. We can observe changes in migration flows as various other economic variables alter across time and space. That, combined with an estimate of the costs of migrating, then allows us to consider both levels and changes in BCRs over time (for instance, possible reversals of flow as benefit patterns changes over space, reducing the BCRs for moving to some locations while increasing the BCRs for other possible moves).

Another approach, also in the revealed preference vein, is to observe what households spend increased funds on if funds are made available to households with no strings attached. In addition, one can compare results that occur if control over the funds is given to different members of the household. This allows us to see what choices the households make on the true margin (rather than calculating them in the household expenditure survey approach as outlined above) as well as how these choices may vary depending on who controls the funds.

Given this range of possible ways of generating a more robust set of BCRs, it seems hard to argue that the only acceptable standard is highly controlled social experimentation. One is loath to reject entire alternative but generally accepted methods such as multiple regression analysis on panel data sets, as well as carefully considered case studies, in favor of holding to a much smaller set of BCR calculations.

The Attribution Problem

The fundamental question of when there is causality vs. when there is simply correlation is very difficult to untangle for gender issues. While there are occasionally times where the timing pattern of a social innovation such as a law passage can be marked, even in those situations it may be argued that the time was right for such a passage, and that social change was leading to it rather than vice versa. This is a difficult problem for Clots-Figueras to resolve, because she wants to argue that there are some changes that are (apparently necessary) preconditions for other changes to occur. She is particularly interested in making the case that women need to receive full political rights in order for other positive economic changes to occur (of the type that will make it possible for BCRs to be high for subsequent social policy changes). But are such legal changes actually necessary for women to see, for example, improved economic participation and improved health outcomes?

As part of a very interesting research agenda that exploits US state-level differences in law passage in the nineteenth and twentieth centuries, Geddes and Lueck (2002) argue that increases in wealth and the growth of cities are associated with the expansion of women's rights—in other words, that economic development (and the accompanying urbanization) led to more rights for women, not vice versa. On the other hand, Geddes, Lueck, and Tennyson (2012) also show, using historical data for the US by state between 1850 and 1920, that "expanding women's economic rights resulted in higher relative rates of school attendance by girls."

It certainly appears plausible, particularly in many countries where women still lag well behind men in terms of having land ownership rights, inheritance rights, and legal protection for equal treatment, that these changes are indeed necessary preconditions for equal economic outcomes by gender. But it may ben harder to argue that political representation quotas are necessary preconditions as well, particularly given that most currently highly economically developed countries reached that condition without the use of such quotas. In addition, one might suspect quotas mainly affect distribution rather than efficiency. Indeed, Chattopadhyay and Duflo (2004) indicate that political reservations for women on village councils in India had the effect of increasing investments in infrastructure that were more relevant to the needs of their own gender, but could not be said to be clearly more or less efficient than the situation of no reservations.

The causality problem continues with regards to some of the other topics that Clots-Figueras discusses in her paper. Is violence against women most cost-effectively reduced through direct action against violence, or through improving women's economic standing so that they can themselves stand up against violence, including having the economic wherewithal to leave domestic violence situations? Early marriage is another interesting case. We see a decline in early marriage (indeed, a move to very late marriage) as societies undergo economic development; is it cost-effective to ban such marriages, or allow them to become less attractive options over time as women's value in the labor market rises, particularly if they can complete their schooling before marriage?

Interestingly, in her very influential survey article, Duflo (2012) argues that both directions hold: economic development reduces gender inequality, but empowering women also benefits development. However focusing on just one or the other causal linkage does not necessarily work; she argues that "the interrelationships are probably too weak to be self-sustaining, and that continuous policy commitment to equality for its own sake may be needed to bring about equality between men and women." While there is no concern voiced here for the BCR of such continuous policy commitment, that would be a natural next question to ask in the context of the Clots-Figueras paper: is it cheaper to support change flowing in one direction, or the other?

The Limits of Positive Analysis

Clots-Figueras has to fall back on arguing that some policy goals are preconditions because they are not readily amenable to BCR analysis, but then cannot make a clear statement

regarding why they are justified; indeed, clear historical precedents not always available regarding the precise causal direction. In the case of her paper, these preconditions cut wide swaths: education for women (but how much, and for all women?); reproductive and maternal health (again, how much health, measured by whatever indicator is chosen, and for all?); equal rights to own and inherit property, sign contracts, own businesses and bank accounts (harder to argue against these, but are they really preconditions?). Overall, pontificating that certain things should be done may get us economists in less trouble if we would actually fall back on normative justifications regarding their inherent morality rather than our using instrumental justifications regarding necessary preconditions to achieve higher economic efficiency (in part because the argument regarding preconditions may not be substantiated empirically).

In addition, Clots-Figueras runs up against the usual limitations of the economist's main tool, positive analysis. Positive economic analysis is, other than in highly circumscribed cases, unable to justify actions/changes based on BCR analysis without falling back on normative grounds (whether acknowledged or implicit). In the case of gender issues, there is generally going to be some redistribution between women and men or between different subgroups of women and men (whether differentiated by marital status, class, caste, race, or ethnicity). Thus the usual iron constraint of conventional economic welfare theory, Pareto optimality, holds here as with practically all policy measures. Arguing that a policy should be done because it has a sufficiently high BCR ignores the question of who bears the costs and who reaps the benefits; accepting any policy with a BCR higher than one is essentially an acceptance of the Kaldor criterion, even if one does not state it thusly.

An alternative approach would be to put forth principles for making choices based on clearly stated philosophical standards (even including the reductionist principle of the Kaldor criterion). Two philosophers of particular relevance to economics are John Rawls (1971) and Robert Nozick (1974) (building on Kant). Under the Rawlsian original position standard, in this context, the question can be posed as to how you would want to structure society if you did not know whether you were going to be born as a man or a woman (as well as not knowing at what social stratum you would be). The implication is likely that you would prefer a system of gender equality and full participation rights for all. Meanwhile, a Nozickian system (building on the Kantian position of treating people as ends rather than as means) would also support gender equality as it has an equal starting point principle and an antienslavement principle; gender inequality, particularly related to disenfranchisement of property rights and voting, is thus insupportable. Similarly, both systems might support expansion of workers' rights for both genders. However, political representation gender quotas appear unsupportable under the Nozickian system and only potentially supportable under the Rawlsian framework.

Conclusion

Clots-Figueras raises much good food for thought in her article and takes an admirably clarifying approach to determining which Post-2015 goals are relevant in a gender context. Her focus on BCRs can be both critiqued and extended, with extensions both philosophical and empirical in nature. While more work remains to be done in calculating and

considering BCRs for gender-related issues, her paper takes a crucial first step in the right direction.

References

R. Chattopadhyay and E. Duflo, 2004: Women as policy makers: Evidence from a randomized policy experiment in India, Econometrica 72(5), 1409-1443.

E. Duflo, 2012: Women empowerment and economic development, Journal of Economic Literature 50(4), 1051-1079.

Geddes, R. and D. Lueck, 2002: The gains from self-ownership and the expansion of women's rights, American Economic Review 92(4), 1079-1092

Geddes, R., D. Lueck, and S. Tennyson, 2012: Human capital accumulation and the expansion of women's economic rights, Journal of Law and Economics 55(3), 839-867.

Gregory, R.G., R. Anstie, A. Daly, and V. Ho, 1989: Women's pay in Australia, Great Britain, and the United States: The role of laws, regulations, and human capital. In Pay Equity: Empirical Inquiries, R.T. Michael and H.I. Hartmann (eds.), 222-246. Washington, DC: National Academy Press.

International Monetary Fund, 2014: International financial statistics.

R. Nozick, 1974: Anarchy, State, and Utopia. New York: Basic Books.

J. Rawls, 1971: A Theory of Justice. Cambridge, Mass: Belknap.

United Nations Statistics Division, 2014: Statistical databases.

World Bank, 2014: World development indicators.

World Economic Forum, 2013: The global gender gap report 2013.

This paper was written by Joyce P. Jacobsen, Professor of Economics at Wesleyan University. The project brings together 62 teams of economists with NGOs, international agencies and businesses to identify the goals with the greatest benefit-to-cost ratio for the next set of UN development goals.

For more information visit post2015consensus.com

COPENHAGEN CONSENSUS CENTER

Copenhagen Consensus Center is a think tank that investigates and publishes the best policies and investment opportunities based on how much social good (measured in dollars, but also incorporating e.g. welfare, health and environmental protection) for every dollar spent. The Copenhagen Consensus was conceived to address a fundamental, but overlooked topic in international development: In a world with limited budgets and attention spans, we need to find effective ways to do the most good for the most people. The Copenhagen Consensus works with 100+ of the world's top economists including 7 Nobel Laureates to prioritize solutions to the world's biggest problems, on the basis of data and cost-benefit analysis.