CENTER FOR
ECONOMIC
BEHAVIOR &
INEQUALITY

CEBI WORKING PAPER SERIES

Working Paper 06/24

IDENTIFICATION OF MARGINAL TREATMENT EFFECTS USING SUBJECTIVE EXPECTATIONS

Joseph Briggs

Andrew Caplin

Søren Leth-Petersen

Christopher Tonetti

ISSN 2596-447X

CEBI
Department of Economics
University of Copenhagen
www.cebi.ku.dk

Identification of Marginal Treatment Effects using Subjective Expectations

Joseph Briggs
Goldman Sachs

Andrew Caplin
New York University & NBER

Søren Leth-Petersen
University of Copenhagen, CEBI & CEPR

Christopher Tonetti*
Stanford GSB & NBER

March 2024

We develop a method to identify the individual latent propensity to select into treatment and marginal treatment effects. Identification is achieved with survey data on individuals' subjective expectations of their treatment propensity and of their treatment-contingent outcomes. We use the method to study how child birth affects female labor supply in Denmark. We find limited latent heterogeneity and large short-term effects that vanish by 18 months after birth. We support the validity of the identifying assumptions in this context by using administrative data to show that the average treatment effect on the treated computed using our method and traditional event-study methods are nearly equal. Finally, we study the effects of counterfactual changes to child care cost and quality on female labor supply.

JEL Codes: C32, C52, C83, J13, J22

*Caplin thanks the Nomis and the Sloan Foundation for support. Leth-Petersen is grateful for financial support from the Independent Research Fund Denmark and CEBI. Center for Economic Behavior and Inequality (CEBI) is a center of excellence at the University of Copenhagen financed by grant DNRF134 from the Danish National Research Foundation. We have benefited from discussions with Julie Cullen, Gordon Dahl, Rebecca Diamond, Matthew Gentzkow, Lihua Lei, Magne Mogstad, David Ritzwoller, Brad Ross, and Alex Torgovitsky. We thank seminar participants at many universities, conferences, and research centers for useful comments.

1 Introduction

Estimating the causal effect of being treated by a policy on economic outcomes is a central problem in applied economics. This task is often complicated by the possibility that the effect of treatment varies across individuals and that selection into treatment is not random, but rather driven by unobserved factors that also affect outcomes. As a result, quasi-experimental methods are used to estimate treatment effects, but such estimates are potentially specific to the subpopulation for whom the quasi-experiment generates variation. Assessing the external validity of treatment effect estimates, i.e., the degree to which estimates can be extrapolated to other subpopulations, is a key challenge in applied econometric analyses.

Marginal treatment effects ($MTEs$) can be used to address issues that arise due to nonrandom selection into treatment. MTEs are essentially the average treatment effect for groups of individuals who have the same latent propensity to be treated.[1] MTEs are useful because they can be combined with different weights to calculate any average treatment effect of interest, such as the average treatment effect on the treated.[2] Even if individual treatment effects (ITE) were known for all individuals, it would still be necessary to know who would select into treatment in order to assess the effect of a policy and to calculate many treatment effects of interest. Identifying MTEs is desirable because they precisely combine ITEs and treatment propensities to achieve these goals.

In practice, identifying MTEs is challenging. Identification of $MTEs$ for the full population using instrumental variables requires instruments that are often unavailable, i.e., IVs that generate variation in treatment for individuals across the full support of the distribution of the latent propensity to be treated. Because such instruments are rare, there is an active literature that aims to overcome this identification challenge by using available instruments and imposing parametric assumptions on the relationship between treatment propensity and outcomes. We develop an alternative, nonparametric, identification strategy that uses survey data on subjective expectations instead of instruments.

We show that subjective expectations data with information about choice probabilities identify MTEs.[3] The key insight to our approach is that respondents often have superior information than the econometrician about their likelihood of selecting into treatment. This is the case when individuals are likely to have a high degree of private information about selection into treatment and/or when decisions are made under uncertainty and individuals

[1] See, e.g., Heckman and Vytlacil (1999), Heckman and Vytlacil (2005), Heckman and Vytlacil (2007a), and Eisenhauer, Heckman, and Vytlacil (2015).

[2] Tables IA and IB in Heckman and Vytlacil (2005) provide the specification and weights to compute various average treatment effects using $MTEs$.

[3] We build on foundational work that uses surveys to measure subjective beliefs, as summarized by Manski (2004).

behave based on their beliefs about the outcome of treatment.[4]

Our main methodological contribution is to state assumptions under which survey data of subjective expectations are sufficient to identify the latent propensity to select into treatment and the marginal effect of treatment. First, we show that the subjective probability of being treated identifies the latent determinant of selection into treatment if people with the same observable characteristics use a common model to forecast their selection into treatment. We then show that, under this same assumption, subjective treatment probabilities and subjective expectations about treatment-contingent outcomes identify subjective marginal treatment effects ($seaMTE$s).

Another contribution is to explore the relationship between subjective and objective MTEs. When the individuals' subjective expected outcomes equal objective expected outcomes, then the subjective marginal treatment effect equals the objective one. Thus, it is sufficient, but not necessary, for individuals to have rational expectations. We also show that our method is robust to common types of errors in the measurement of subjective expectations (e.g., rounding or optimism) as long as a rank-preserving condition is satisfied. In addition to identifying MTEs, another useful feature of our approach is that beliefs can be elicited in hypothetical policy scenarios to perform counterfactual analysis.

As a proof-of-concept, we use the method to study how the treatment of child birth affects female labor supply. We focus on female labor supply around child birth because it is an important choice which individuals have likely considered thoughtfully and are likely to have a high degree of private information regarding selection into treatment. We estimate $seaMTEs$ using a custom-designed Danish survey that collects data on subjective beliefs about the likelihood of childbirth in the near term and birth-contingent labor supply for a representative sample of about 11,000 Danish women aged 18–40. We show that fertility beliefs align with realized historical fertility in the Danish population and that beliefs about the effect of having a child on labor supply stated in the survey align with effects estimated on administrative data using an event study design. The results indicate that $MTEs$ are roughly constant across levels of fertility beliefs. These findings suggest that, after conditioning on typical demographic and economic observables, there is little selection into having children based on an individual-specific latent propensity to work after having a child. Having a child reduces the likelihood of working after birth by about 50 percent 9 months later, around 3 percent 18 months later, and around 0 percent 36 months later.

In addition to stating sufficient conditions for identification and estimating $seaMTE$s

[4]There is increasing evidence that subjective beliefs matter for choice behavior, e.g., Arcidiacono, Hotz, Maurel, and Romano (2020), Coibion, Gorodnichenko, and Ropele (2019), Delavande (2008), Delavande and Zafar (2019), Lochner (2007), Wiswall and Zafar (2020), Colarieti, Mei, and Stantcheva (2024).

in the labor-supply context, we also propose and implement a test that is informative of whether subjective and objective MTEs are equal. In our empirical context, we show that $seaMTE$ and MTE are likely to be equal. We compute average treatment effects on the treated (ATTs) using our estimated $seaMTE$s and show they are essentially equal to ATTs computed from administrative data with a traditional event-study method. While this could be a pure coincidence, we view this as evidence that $seaMTE = MTE$ in this context. More generally, we view this result as a proof of concept that our proposed methodology can work. Even if $seaMTEs$ are only valuable if they equal objective MTEs and we can only tell if they are equal if we have a valid traditional estimator for some average treatment effect, there is still value in estimating $seaMTEs$, because the $seaMTE$ allows us to compute any average treatment effect of interest.

Furthermore, $seaMTE$s are useful because we can use them to predict the outcome of never before seen policies, properly controlling for how selection into treatment will be affected by the considered policies. In the Danish survey we also elicit subjective childbirth and state-contingent labor supply probabilities under hypothetical childcare policy regimes with different cost and quality of care. We find that these policy changes have significant effects on expected fertility. We then show that MTE functions are unchanged in the free childcare regime, but that $MTEs$ shift significantly (reflecting reduced labor supply) if childcare were to be substantially more expensive and lower quality. These results highlight the value of the survey methodology: making childcare free falls within the class of policy changes for which policy-relevant treatment effects ($PRTEs$) constructed from baseline $MTEs$ are valid, whereas making childcare much more expensive and lower-quality is a policy for which $PRTEs$ constructed from baseline $MTEs$ are not valid. Despite the fact that the latter policy change cannot be inferred from baseline $MTEs$, our data provide insight into how such a policy change might affect the level and shape of $MTEs$ and who might respond most to the policy. For example, our estimates suggest that lowering quality and tripling the price of child care in Denmark would lead to a drop in fertility of about 8 percentage points and to a reduction in labor supply among women who have children of about 30 percentage points 36 months after birth.

The rest of the paper is organized as follows. Next, we describe how our paper relates to existing literature. In Section 2, we introduce the Generalized Roy model of treatment and outcomes and develop our identification strategy by mapping subjective probability distributions to model parameters. In Section 3, we present an application to female labor supply after childbirth based on a custom Danish survey. Section 4 concludes.

1.1 Relationship to Literature

Treatment effect heterogeneity and nonrandom selection into treatment have are generally recognized as posing a significant challenge to the interpretation of estimated treatment effects. Questions regarding external validity arise in almost all empirical settings. We contribute to a body of research by Heckman and Vytlacil (1999, 2001, 2005, 2007b) that develops and uses $MTEs$ to study unobserved heterogeneity and the external validity of estimates. Heckman and Vytlacil (1999) shows that $MTEs$ can be estimated by the method of local instrumental variables. However, nonparametric identification of $MTEs$ for the full population requires a continuous instrument that generates a propensity score with full support conditional on all combinations of conditioning variables. This requirement is quite stringent, and in practice extra assumptions are often invoked to gain identification.

One approach is to use point-identified treatment effects and either assume no treatment effect heterogeneity (see, e.g., Heckman and Robb (1985), Angrist and Fernandez-Val (2013)) or to parameterize unobserved heterogeneity. Prominent examples include assuming normality of idiosyncratic selection and outcome components (Bjorklund and Moffitt (1987), Heckman, Tobias, and Vytlacil (2003)), monotonicity and concavity (Manski and Pepper (2000)), full independence assumptions which ensure that the MTE slope is independent of conditioning variables X (see, e.g., Aakvik, Heckman, and Vytlacil (2005); Carneiro, Heckman, and Vytlacil (2011); Carneiro, Lokshin, and Umapathi (2017)), and additive separability (Brinch, Mogstad, and Wiswall (2017)).

A second approach is to bound certain treatment effect parameters (see, e.g., Manski (1990), Manski (1997), Manski (2003), Heckman and Vytlacil (2007b)) based on parameters identified from observational data. A recent advance in this area is Mogstad, Santos, and Torgovitsky (2018), which exploits information from a wide range of identified parameters, termed marginal treatment responses (MTR), to infer nonparametric bounds on $MTEs$ that are not point-identified in observational data.

Our method differs from this previous work in that it nonparametrically point-identifies $MTEs$ over the entire support of the latent determinant of selection into treatment using subjective expectations data. This approach allows us to study the external validity of LATE and other treatment effect parameters and to identify treatment effects for subpopulations that lack sufficient observational data on treatment status, outcomes, and instruments.

Our paper also relates to a body of research on subjective expectations. Following Manski (2004)'s influential survey of early work in this area, a number of studies measure subjective expectations and use them to better understand economic choices and outcomes.[5] We most closely relate to papers that elicit state-contingent outcomes. Quantification of causal treatment effects involves contrasting treatment-contingent outcomes while holding factors that affect outcomes other than the treatment constant. It is by nature impossible to make such within-person comparisons with data on realized outcomes. This is called the fundamental evaluation problem in the treatment effect literature (Rubin, 1974). Collecting data about subjective state-contingent outcomes explicitly addresses this issue.

Asking about state-contingent expectations dates back to at least Dominitz and Manski (1996) who, in an exploratory study, asked a group of high school students and college undergraduates about probabilistic expectations in alternative scenarios for future schooling. A number of recent studies also use subjective expectations of state-contingent outcomes to model educational choices (Arcidiacono, Hotz, and Kang (2012); Zafar (2011); Zafar (2013); Stinebrickner and Stinebrickner (2013); Wiswall and Zafar (2015); Kaufmann (2014); Attanasio and Kaufmann (2014); Attanasio and Kaufmann (2017); Delavande and Zafar (2019)). Closely related to our study is the Arcidiacono, Hotz, Maurel, and Romano (2020) paper that studies occupational choice using data for 173 Duke undergraduates with information about occupation-specific earnings beliefs, probabilities of choosing these occupations, and realizations. They show how subjective expectations data can be used to recover ex ante individual treatment effects of occupational choice on earnings and how individual treatment effects relate to the ex ante average treatment effect and the ex ante average treatment effect on the treated. A similar approach is applied by Giustinelli and Shapiro (2024) who collect data on beliefs about health status and retirement and show that such data can be used to estimate individual treatment effects of health on labor supply. See also Hudomiet, Hurd, Parker, and Rohwedder (2021), who use individuals' subjective conditional probability of working at age 70 to estimate the causal effect of job characteristics on retirement.

Relative to these studies, we show how and under what assumptions individuals' subjective expectations of the treatment propensity and of state-contingent outcomes identify marginal treatment effects and, by extension, any conventional average treatment effect parameter. That is, in addition to the distribution of individual treatment effects that other studies measure, we are concerned with nonrandom selection into treatment. We also specify the conditions under which historical data can be used to validate treatment effects

[5] See, e.g., Pistaferri (2001), Attanasio (2009), Hurd (2009), Delavande and Rohwedder (2011), van der Klaauw (2012), Delavande (2014), Delavande and Kohler (2016), Hendren (2017), Giustinelli and Manski (2018), Alesina, Stantcheva, and Teso (2018), Bailey, Davila, Kuchler, and Stroebel (2019), Kuchler and Zafar (2019), Christelis, Georgarakos, Jappelli, and van Rooij (2020), Wiswall and Zafar (2020), and Andersen and Leth-Petersen (2021). Also using survey expectations, Jappelli and Pistaferri (2014) and Fuster, Kaplan, and Zafar (2021) investigate the effect of unobserved state-contingent choices in the context of spending responses to fiscal policy.

estimated using subjective expectations by linking expectations data to historical administrative records. We view this as a useful alternative to waiting until choices and outcomes have been realized, as is done by Arcidiacono, Hotz, Maurel, and Romano (2020) and Wiswall and Zafar (2020) in their studies of human capital investment and expectations about family and career. Moreover, we provide a method to infer treatment effects in counterfactual policy environments that provide a suggestive test for the policy invariance of $MTEs$ estimated from historical data.

Finally, to our knowledge, we are the first study to document the role of unobserved heterogeneity in the effect of childbirth on labor supply. Within this large literature, our study relates in particular to two recent papers that investigate female labor supply following child birth in Denmark. Kleven, Landais, and Sogaard (2019) use an event study design and annual administrative data for the period 1985-2003 and find that female labor supply is reduced after child birth by about 12–14 percent. Lundborg, Plug, and Rasmussen (2017) exploits randomness in the success of in vitro fertilization (IVF) treatments in Denmark to construct IV estimates of the effect for women with no prior children. They find that IVF treated women who have children, on average, reduce labor supply by about seven percent compared to IVF treated women who do not have children. While IVF treated women are a selected group, these estimates are in the same ball park as those presented by Kleven, Landais, and Sogaard (2019). We show that selection on unobserved characteristics is not likely to be important in the Danish context thereby tying together the findings of Kleven, Landais, and Sogaard (2019) and Lundborg, Plug, and Rasmussen (2017).

2 A Survey Method to Identify Marginal Treatment Effects

In this section we first describe the Generalized Roy model (Roy, 1951) that we take as the data generating process and then introduce our method of identifying $MTEs$ using subjective expectations. We first present the model in abstract and then provide an interpretation of the variables in the context of female labor supply and childbirth.

2.1 The Generalized Roy Model

Let $D^i \in \{0,1\}$ be an indicator for whether individual i is subject to a binary treatment, with 1 denoting treatment. Let $Y^i_D \in \mathbb{R}$ be a random variable that is the outcome for individual i if $D^i = D$. Selection into treatment D^i is determined by the difference of the random variable

μ^i and an idiosyncratic component $V^i \in \mathbb{R}$ (i.e., an idiosyncratic distaste for treatment):

$$D^i = \begin{cases} 1 & \text{if } \mu^i \geq V^i \\ 0 & \text{if } \mu^i < V^i. \end{cases} \tag{1}$$

The timing of the model is as follows and is structured to interface with the timing of events and surveys. By $t = 0$ (the time of the survey), individual constants X^i and V^i are determined. At $t = 1$, random variable μ^i is drawn from CDF $K_{\mu|X^i}$ that is a function of variables $X \in \mathbb{R}^{d_x}$. Treatment status D^i is determined in period $t = 1$ as a function of μ^i. At $t = 2$ the random variable ϵ^i is drawn from CDF $J_{\epsilon|X^i}$ that is a function of variables $X \in \mathbb{R}^{d_x}$, with $\mu^i \perp \epsilon^i | X^i$. The outcome Y_D^i is determined in period $t = 2$ as a function of ϵ^i. Crucially, the outcome is modeled as a function of treatment status (D^i), individual characteristics (X^i, V^i) and shocks (μ^i, ϵ^i): $Y_D^i(X^i, V^i, \mu^i, \epsilon^i)$.

Our selection equation is a generalized version of the standard selection framework (see, e.g., Heckman and Vytlacil (2005)). Both V and X are individual characteristics that affect selection into treatment. As usual, the key difference is that X is observable to the econometrician, while V is not. We model the first term in the selection equation (μ) as a stochastic function. Since we often have data on individuals before selection into treatment occurs and selection into treatment is often not perfectly forecastable by the individual, the stochastic case is relevant for many applications.

Information, Beliefs, and Behavior. At $t = 0$, before selection into treatment occurs, the individual knows X^i, V^i, but does not know μ^i or ϵ^i (and thus does not know D^i or Y_D^i).

At $t = 0$, let G^i denote individual i's subjective probability of being treated:

$$G^i := P^i(D^i = 1). \tag{2}$$

If the individual has rational expectations, i.e., if their subjective beliefs align with the probabilities dictated by the model, then

$$G^i = 1 - K_{\mu|X^i}(V^i). \tag{3}$$

At $t = 0$, let h_D^i denote individual i's subjective probability density function over treatment-contingent outcome Y_D^i. Let $\mathbb{E}^i\left[Y_D^i\right]$ be the treatment-contingent subjective expected value of the outcome. Finally, let $seaITE^i$ be the subjective ex ante individual treatment effect, i.e., the subjective expected difference in outcomes across treatment states.

$$P^i(a < Y_D^i < b) =: \int_a^b h_D^i(y) dy, \tag{4}$$

$$\mathbb{E}^i \left[Y_D^i\right] := \int h_D^i(y)\ y\ dy, \tag{5}$$

$$seaITE^i := \mathbb{E}^i \left[Y_1^i\right] - \mathbb{E}^i \left[Y_0^i\right]. \tag{6}$$

The starting point of our analysis is that individuals' preferences and subjective beliefs affect their behavior. The model captures this by allowing the V^i that determines selection into treatment to be a function of the subjective probability of outcomes ($h_D^i(\cdot)$), along with other unspecified idiosyncratic components like preferences and costs. Furthermore, state-contingent outcomes are a function of V^i, thus any statistic that is a function of differences in state-contingent outcomes—such as MTEs—will be a function of these subjective beliefs.

Selection into treatment is determined by V^i and the realization of the random variable μ^i, not by the beliefs about selection into treatment G^i. Subjective beliefs about selection into treatment affect individuals' answers to our survey questions at $t = 0$, but not any real-world behavior like treatment status or outcomes. We will discuss below how deviations from rational expectations about the distribution of μ^i and ϵ^i affect our estimates of various treatment effects.

Example: Female Labor Supply and Childbirth through the Lens of the Model. To make this more concrete, let's consider a woman who may or may not give birth to a child 24 months in the future and who may or may not be working 36 months in the future, where the binary treatment is having the child or not and the outcome is working or not. There are individual characteristics in our data (X^i) like age, income, wealth, and partner status, that may affect her propensity to have a child, which is captured by the distribution of μ^i depending on X^i. She may or may not love being around children and she may or may not have access to low-cost childcare which may affect her desire to have a child but are not measured in our data; such preferences and costs would be captured by V^i. She may also have beliefs about how her employment outcomes would be affected by childbirth ($h_D^i(\cdot)$) that can influence her decision to select into having a child, which is also captured by V^i. She may have a chance at being promoted at her job, and whether she gets the promotion may affect her desire to have a child, which would be captured in the stochastic component of μ^i. Furthermore, her desire to work conditional on having or not having a child might depend on whether she got the promotion, which is why the outcome function $Y_D^i(X^i, V^i, \mu^i, \epsilon^i)$ depends directly on μ^i, and not just on D. Her desire to work after having a child may also be a function of whether she loves being around children, which is captured by the outcome function $Y^i_D(X^i, V^i, \mu^i, \epsilon^i)$ depending directly on V^i. Finally, there can be some event that affects her desire to be employed that is independent of child status and promotion status, like a parental illness, that would be captured in ϵ^i.

Normalization and Notation. As is common in the literature, we assume V is continuous.[6] Similarly, we assume G is continuous.

Assumption 1. *V and G are continuously distributed, conditional on X.*

This assumption allows us to normalize arguments in selection equation (1) to the unit interval. Define the CDF of V conditional on X as $F_{V|X}$. Let $M^i = F_{V|X^i}(\mu^i)$ and $U^i = F_{V|X^i}(V^i)$. U^i denotes individual i's order in the distribution of latent selection component V, conditional on X. M^i is the fraction of people who have $V < \mu^i$, conditional on X. Since $F_{V|X}$ is an increasing function, $M^i \geq U^i \iff \mu^i \geq V^i$. Thus Assumption 1 allows equation (1) to be rewritten in terms of normalized variables as

$$D^i = \begin{cases} 1 & \text{if } M^i \geq U^i \\ 0 & \text{if } M^i < U^i. \end{cases} \tag{7}$$

Let $N(u, x) := \{i : U^i = u, X^i = x\}$ and $N(x) := \{i : X^i = x\}$. $N(u, x)$ is the set of individuals who have the same characteristics ($X^i = x$) and latent propensity to select into treatment ($U^i = u$). Let $\mathbb{I}(i \in N(u, x))$ be an indicator function that equals 1 if individual i is in $N(u, x)$ and equals zero otherwise.

Similarly, define the CDF of G conditional on X as $F_{G|X}$ and $\tilde{U}^i = 1 - F_{G|X^i}(G^i)$. Then $\tilde{U}^i$ is individual i's order in the distribution of the subjective probability of selecting into treatment, conditional on X, when ordered from highest to lowest. Also similarly define $\tilde{N}(u, x) := \{i : \tilde{U}^i = u, X^i = x\}$.

2.2 Definition of MTEs

Heckman and Vytlacil (1999) and Heckman and Vytlacil (2005) define the ex ante MTE as the expected gain from treatment $(Y^i_1 - Y^i_0)$ conditional on X and U. Below we define two versions of $MTEs$: the ex ante MTE and the subjective ex ante MTE.

[6] See, e.g., Assumption 3 in Vytlacil (2002).

Definition of Marginal Treatment Effects.

$$MTE(u,x) := \mathbb{E}\left[Y_1^i - Y_0^i | U^i = u, X^i = x\right], \tag{8}$$
$$= \frac{1}{|N(u,x)|} \int\int\int \left[Y_1^i(X^i, V^i, \mu^i, \epsilon^i) - Y_0^i(X^i, V^i, \mu^i, \epsilon^i)\right] \, dK_{\mu|x}(\mu^i) \, dJ_{\epsilon|x}(\epsilon^i) \, \mathbb{I}(i \in N(u,x)) \, di.$$

The MTE is essentially the local average treatment effect ($LATE$), local both to members of the population with the same values of X^i and, crucially, to members with a similar latent propensity to select into treatment U^i. We define the ex ante MTE as the expected $LATE$ conditional on the information available before the selection into treatment and outcomes are realized, where the expectation is taken with respect to the probability measures of nature that determine selection and outcomes.

Definition of Subjective Ex Ante MTE (seaMTE). Similarly, we define the subjective ex ante marginal treatment effect ($seaMTE$) as the $LATE$ for individuals with the same values of X^i and $\tilde{U}^i$. The key distinction is that the $seaMTE$ conditions on the subjective treatment propensity ($\tilde{U}^i$ instead of U^i) and uses the subjective probability distributions of the individuals to take expectations ($E^i[\cdot]$ instead of $E[\cdot]$).

$$seaMTE(u,x) := \mathbb{E}\left[\mathbb{E}^i\left[Y_1^i\right] - \mathbb{E}^i\left[Y_0^i\right] | \tilde{U}^i = u, X^i = x\right], \tag{9}$$
$$= \frac{1}{|\tilde{N}(u,x)|} \int\int \left[h_1^i(y)y - h_0^i(y)y\right] \, dy \, \mathbb{I}(i \in \tilde{N}(u,x)) \, di.$$

By definition, if individuals have rational expectations, then $seaMTE(u,x) = MTE(u,x)$. Returning to the example of labor supply after childbirth, if an individual does not have rational expectations about μ, and believes it is more likely that she will get the promotion than is true, then her subjective probability of selecting into treatment may be different from the actual chance, which may lead the $seaMTE \neq MTE$ because $\mathbb{I}(i \in N(u,x))$ may not equal $\mathbb{I}(i \in \tilde{N}(u,x))$. Furthermore, $seaMTE \neq MTE$ is possible if individuals do not have rational expectations about outcomes. For example, Kuziemko, Pan, Shen, and Washington (2018) document a "Mommy Effect" in which women think that they are more likely to work after childbirth than they really are. This can be captured in our framework if, e.g., ϵ captures how well the child tolerates being away from the mother, mothers are overoptimistic about this, and mothers are less likely to work if their child does not tolerate being away from them.

2.3 How to Identify U^i and $seaMTE$ with Survey Data

We now introduce the data available to an econometrician at time zero and specify the conditions under which they can be used to estimate treatment effects. The econometrician observes X but does not observe V. The econometrician also has survey data on individuals' subjective probabilities of selecting into treatment ($\hat{G}^i$) and state-contingent outcomes ($\hat{h}_D^i$), which were measured before treatment selection occurs. Since G^i and h_D^i are both subjective probability distributions, they can be measured directly via appropriately designed surveys. Two ways in which using measured survey responses could lead to inconsistent estimates of $MTEs$ are if people do not have rational expectations and if people do have rational expectations but they cannot report them without error when responding to the survey. For clarity of exposition, we first assume there is zero measurement (survey response) error so that $\hat{G}^i = G^i$ and $\hat{h}_D^i = h_D^i$; we then relax the zero-measurement-error assumption in Section 2.4.

Assumption 2, stated below, provides a condition under which U^i is identified from subjective expectations data.

Assumption 2. $G^i := P^i(D^i = 1) = 1 - \tilde{K}_{\mu|X^i}(V^i)$, *for some CDF* $\tilde{K}_{\mu|X^i}$.

Assumption 2 implies that all individuals who have the same X^i differ in their subjective probability of selecting into treatment only due to differences in their V^i. Furthermore, conditional on X, those with higher V^i have a lower subjective probability of selecting into treatment. $\tilde{K}_{\mu|X^i}(V^i)$ can be thought of as a forecasting function that maps observables X^i into a probability of selecting into treatment, conditional on V^i. Assumption 2 requires that $\tilde{K}$ is common for all individuals who have the same values of X^i, i.e., that people with the same X^i use the same forecasting rule. This assumption is obviously satisfied if individuals have rational expectations, i.e., $\tilde{K}_{\mu|X^i} = K_{\mu|X^i}$, but all that is necessary is the commonality of beliefs.

Assumption 2 delivers our first key result: given measured observables X^i, subjective treatment probabilities, G^i, are sufficient to identify U^i for each individual. The basic idea behind this result is simple. Consider two individuals, i and j, who have the same observable characteristics X. Since they have the same X, they have the same distribution of μ, so they only differ in their expectation of being treated due to their idiosyncratic V^i, V^j. The probability of receiving treatment is decreasing in V and V is in an individual's information set, so individual i will have a higher subjective probability of selection into treatment than individual j iff $V^i < V^j$.

Define the CDF of $\hat{G}$ conditional on X as $F_{\hat{G}|X}$. Define $\hat{U}^i \in [0, 1]$ to be the order of individuals with similar observable characteristics when ranked from highest to lowest by their reported subjective probability of selecting into treatment:

$$\hat{U}^i := 1 - F_{\hat{G}|X^i}(\hat{G}^i), \tag{10}$$

and let, $\hat{N}(u,x) := \{i : \hat{U}^i = u, X^i = x\}$.

Proposition 1. *If Assumptions 1, 2, and zero measurement error hold, then*

$$U^i = \hat{U}^i.$$

Proof. See Appendix A.

Proposition 1 is one of the main results of our paper because it relates a measured object to the key latent parameter that determines selection into treatment. Loosely speaking, measuring the subjective probability of selection into treatment makes the unobserved latent propensity parameter observed, so that we can proceed to estimate MTEs by directly conditioning on the now-observed latent propensity.

Identifying MTEs using Subjective Probabilities of Selection and Expected Outcomes. The individual treatment effect (ITE) is the difference between an individual's treatment-contingent outcomes. ITEs are fundamentally impossible to measure using realized outcomes because individuals can never be simultaneously observed in and out of treatment. With some assumptions, ITEs can, however, be identified using a single cross-section of survey data that measures subjective expectations of outcomes. We are not the first to suggest using subjective expectations about outcomes to estimate ITEs and to use these ITEs to compute average treatment effects.[7] Rather, our contribution is to address nonrandom selection into treatment and compute the MTE by also identifying U^i.

Let $\widehat{seaITE}^i$ be the $seaITE$ computed using the directly elicited state-contingent probability of outcomes from the survey:

$$\begin{aligned} \widehat{seaITE}^i &:= \widehat{\mathbb{E}}^i\left[Y_1^i\right] - \widehat{\mathbb{E}}^i\left[Y_0^i\right], \\ &:= \int \left[\hat{h}_1^i(y)y - \hat{h}_0^i(y)y\right]\, dy. \end{aligned} \tag{11}$$

Similarly, let $\widehat{seaMTE}$ be the $seaMTE$ computed using the elicited $\widehat{seaITEs}$ and subjective

[7]See, e.g., Arcidiacono, Hotz, Maurel, and Romano (2020), Wiswall and Zafar (2020), and Giustinelli and Shapiro (2024) who make a similar point.

treatment propensities:

$$\widehat{seaMTE}(x,u) := \mathbb{E}\left[\widehat{seaITE}^i | X^i = x, \hat{U}^i = u\right]. \tag{12}$$

Proposition 2 presents our result on the identification of $seaMTEs$. The proposition is almost a restatement of the definition of $seaMTEs$, except it explicitly relates to measured data.

Proposition 2. *Let Assumptions 1, 2, and zero measurement error hold. Then*

$$seaMTE(x,u) = \widehat{seaMTE}(x,u).$$

Proof. See Appendix A.

Intuitively, if the latent determinants of treatment are identified, as discussed in Proposition 1, and respondents can accurately report their subjective beliefs regarding state-contingent outcomes, then the $seaMTE$ function is identified solely using data on individual characteristics and reported subjective beliefs.

Corollary 1. *Let Assumption 1 and zero measurement error hold. If individuals have rational expectations, then*

$$MTE(x,u) = \widehat{seaMTE}(x,u).$$

Corollary 1 is true because rational expectations implies $seaMTE(x,u) = MTE(x,u)$ and Proposition 2 delivers $\widehat{seaMTE}(x,u) = seaMTE(x,u)$. Rational expectations are clearly sufficient for subjective to equal objective MTEs, however they are not necessary. For $seaMTE = MTE$, it is sufficient that the subjective means of the state-contingent outcome distributions equal those from the model.

2.4 Measurement (Survey Response) Error

Throughout this section we have assumed that the subjective probability of being treated and subjective state-contingent outcome distributions are measured without error. The key results can be extended to cases with certain classes of measurement error.

For U^i to be identified from $\hat{G}^i$, the reported subjective treatment probabilities need to be linked to the true subjective probabilities in a particular way.

Assumption 3. *Suppose that $\hat{G}^i$ is reported such that the order of individuals equals their order in the true subjective probabilities. That is, $G^i < G^j \iff \hat{G}^i < \hat{G}^j \;\; \forall\, i, j$.*

This is obviously satisfied if reported subjective probabilities equal subjective probabilities (i.e., zero measurement error). Assumption 3 states that the subjective probability of treatment needs be reported such that the order of individuals equals their order in the true subjective probabilities. By definition, when Assumption 3 is satisfied, $\tilde{U}^i = \hat{U}^i$. This assumption allows for there to be systematic bias in survey responses about the probability of selection into treatment, as long as the order of individuals in the distribution is preserved. Second, if responses are only weakly increasing in G, as is the case if respondents round when reporting the subjective probability of treatment, then U^i remains set identified among the group of individuals that report a common probability of receiving treatment. Similarly, set identification is also achieved if D is perfectly forecastable (i.e., μ^i is deterministic), since all individuals would report a treatment probability of one or zero in this case. However, point identification of $U^i \; \forall \; i$ in our method requires that there is sufficient uncertainty regarding μ^i to cause individuals to report $\hat{G}^i$ as an interior probability.

Furthermore, the measured subjective ex ante $ITEs$ in equation (11) are equal to $seaITEs$ in equation (6) if the mean reported subjective state-contingent outcome equals the mean subjective state-contingent outcome, which we will maintain as our Assumption 4. It is obviously sufficient, but not necessary, for subjective belief distributions to be reported without error, i.e., if $\hat{h}^i_D(\cdot) = h^i_D(\cdot)$.

Assumption 4. *For all i and for $D \in \{0,1\}$, for some deterministic function $C^i(y)$ and random variables $\eta^i_D(y)$,*

$$\begin{aligned}&\hat{h}^i_D(y) - h^i_D(y) = C^i(y) + \eta^i_D(y),\\ &\mathbb{E}\left[\eta^i_D(y)\right] = 0 \; \forall \; y.\end{aligned}$$

Proposition 2 assumed zero measurement error. This assumption can be relaxed to allow for two classes of measurement error. First, if subjective beliefs regarding state-contingent outcomes are reported with some deterministic error that is common within individual across treatment states, then $\widehat{seaITE}^i = seaITE^i$ and, thus, $\widehat{seaMTE} = seaMTE$. Second, even if $\widehat{seaITE}^i \neq seaITE^i$, it is still possible for $\widehat{seaMTE} = seaMTE$. If the error in reported beliefs has a random component that differs within individuals across treatment status, the subjective ex ante $MTEs$ are still identified. Some $\widehat{seaITE}^i$ will be greater than $seaITE^i$ and some less than, but averaging the mean-zero errors across individuals will still give that $\widehat{seaMTE} = seaMTE$.

In summary, Proposition 1 can be rewritten to relax the assumption of zero measurement error by instead including Assumption 3. Proposition 2 can relax the zero-measurement-error assumption by including Assumptions 3 and 4. Proposition 1' implies Proposition 1, Proposition 2' implies Proposition 2, and Corollary 1' implies Corollary 1. See Appendix A for further discussion of measurement error.

Proposition 1'. *If Assumptions 1, 2, and 3 hold, then* $U^i = \tilde{U}^i = \hat{U}^i$.

Proposition 2'. *If Assumptions 1, 2, 3, and 4 hold, then* $seaMTE(x, u) = \widehat{seaMTE}(x, u)$.

Corollary 1'. *Let Assumptions 1, 3, and 4 hold. If individuals have rational expectations, then* $MTE(x, u) = \widehat{seaMTE}(x, u)$.

3 Application: The Effect of Childbirth on Female Labor Supply in Denmark

In this section we apply our method to study the effect of childbirth on labor supply. We analyze data from a Danish custom survey in which we elicit subjective beliefs concerning future child birth and work. This Danish survey showcases how the survey methodology can ideally be applied to quantify the objects needed for estimating subjective ex ante $MTEs$ as described in Propositions 1 and 2. Furthermore, survey responses from the Danish survey can be matched to administrative records which allows us to assess the representativeness of the sample as well as validate survey responses against historical behavior.

We begin with presenting the institutional context in Denmark and the Danish custom survey. Next, we present the sampling and show how to compute $MTEs$, as described in Propositions 1 and 2. We then compare the survey data on expected fertility and labor supply to historical fertility and labor supply data from administrative registries. Finally, we use answers from questions about expected fertility and labor supply in counterfactual policy scenarios.

3.1 Institutional Context: Child Care and Parental Leave in Denmark

In Denmark the labor participation rate of working-age females is about 80 percent. This is facilitated by a well developed and heavily subsidized high-quality public child care system and maternity leave policies that protect employment. Public day care is organized by municipalities and it is the dominant type of day care provided. All children are eligible for day care except if a parent is on publicly funded maternity or parental leave. Opening hours for public day care are typically 6:30am to 5:00pm on weekdays, thus covering the entire work day throughout the work week for most people. Public day care is regulated by law and the objective is to ensure child care availability for all children and to secure child development and learning. To ensure that public day care in Denmark is of high quality, there are rules that regulate the type and quantity of staffing, physical surroundings, safety, and hygienic standards. Furthermore, municipalities are required by law to monitor child care institutions to ensure that the personnel have sufficient qualifications and that safety and hygienic standards are met. An important dimension of quality is the staffing. In nurseries (0-2 years) there are typically three children per employee and in kindergartens (3-5 years) there are typically six children per employee and about half of the employees have formal pedagogical training. Public day care is heavily subsidized such that users typically pay about one third of the total cost. The typical monthly user cost for a nursery slot is about 3,000DKK (1USD $\approx$ 7DKK) and 1,700DKK for a kindergarten slot. The design of the public day care system has been in place since the 1970s, which means that child care opportunities are well known and predictable. 83 percent of all children aged 0-5 years are enrolled in public or publicly funded day care.

Danish parents are entitled to up to 52 weeks of leave in connection with having a child. 18 weeks, starting four weeks before child birth, are earmarked to the mother and two weeks are earmarked to the father. The remaining 32 weeks can be shared between parents as they prefer. The parental leave scheme guarantees job security and a minimum income corresponding to the unemployment insurance level. For a full time employee this amounts to 90 percent of earnings and a maximum of about 19,000DKK per month before taxes in 2020. These rules have been in place since 2002. We refer to Andersen (2018) for a more detailed description of the history of the parental leave system.

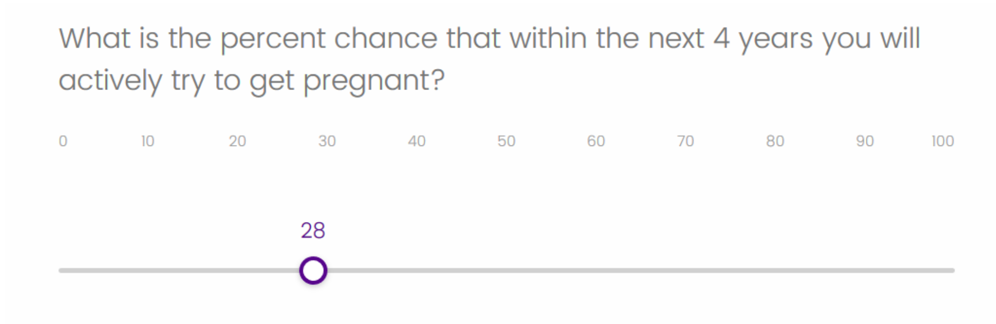
3.2 Data Collection in Denmark

3.2.1 The Custom Survey

The core survey instruments collect the subjective probability of having a child within the next 4 years (i.e., selecting into treatment) and the probabilities of working at horizons $s = -12, -3, 3, 9, 18,$ and 36 months (i.e., distributions over outcomes) after having the child. We measure both treatment probability and outcome distributions by asking respondents to report their beliefs conditional on being in certain states or taking certain actions. Combined, these questions provide the key data objects defined in Section 2, i.e., state-contingent subjective expected outcomes, $\widehat{\mathbb{E}}^i \left[Y^i_{1,s}\right]$ and $\widehat{\mathbb{E}}^i \left[Y^i_{0,s}\right]$, as well as the subjective probability of being treated, $\hat{G}^i$.

To measure fertility, we first ask respondents how likely it is that within the next 4 years they are going to try to have a child. We then ask, conditional on trying and not trying, how likely they are to have a child. The following questions are shown sequentially:

Figure 1: Response Slider.



- What is the percent chance that, within the next 4 years, you will actively try to get pregnant?

- Suppose that, within the next 4 years, you **do** actively try to get pregnant. What is the percent chance that you will have a child in the next 4 years?

- Suppose that, within the next 4 years, you **do not** actively try to get pregnant. What is the percent chance that you will have a child in the next 4 years?

Following each question, respondents are shown a slider with a random initial starting point (respondents had previously been presented with a screen that described how this response technology functioned). As respondents slide left and right to choose the probability, the slider updates the selected number so respondents can see their selected percent chance, as shown in Figure 1.

Because our outcome variable, female labor supply at different horizons, is Bernoulli, collecting the subjective state-contingent outcome distribution entails collecting the probability of working if respondents do and do not have a child at different horizons. We again utilize conditional probabilities to collect these distributions

- Suppose that you do have a child in the next 4 years. What is the percent chance that your partner will take parental leave of at least two months in the first year after the child is born?

- Suppose you do have a child in the next 4 years and your partner **does** take at least two months parental leave. In this scenario, what is the percent chance you will be working exactly [[12/3] months before/[3/9/18/36] months after] your child is born?

- Suppose you do have a child in the next 4 years and your partner **does not** take at least two months parental leave. In this scenario, what is the percent chance you will be working exactly [[12/3] months before/[3/9/18/36] months after] your child is born?

Respondents report their answers to the second and third questions listed above in a set of stacked sliders, where each line corresponds to a different horizon. Finally, we also collect the probability that respondents will be working if they do not have a child:

- Suppose that you do not have a child anytime in the next 4 years. Please think of your work situation RANDOMIZE[.5/1/.../3.5/4] years (RANDOMIZE[6/12/.../42/48] months) from now. What is the percent chance that you will be working during this month?

3.2.2 Sampling and Link to Registry Data

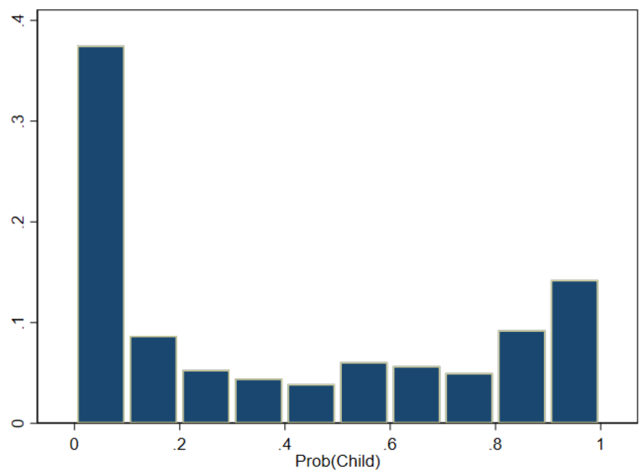
The sample invited to participate in the Danish survey is recruited from the population registry. The population registry is a complete registry of all persons who are born or have ever had a registered address in Denmark. It contains the CPR-number, which is a personal identifier applied universally to record any contact an individual has with the public sector. For the purpose of conducting our survey we have been given access to a random sample of women aged 18-40 on December 31, 2018. The survey was fielded in May 2019. Invitations to participate were sent out using an official email account called *e-boks* which all Danes are equipped with. 51,000 emails were sent out, and we collected 11,090 responses.

In addition to the survey data we also use third-party reported administrative register data compiled by Statistics Denmark from various government agencies. In particular, we use the registries that document household composition, including how many children women have and when they were born. We also rely heavily on data from an administrative register, called the *e-income* registry, that contains employer-reported monthly wages and salaries for employees. All firms in Denmark have to report earnings and hours supplied for each employee, and hence whether the person is working, to the tax agency (SKAT). Since tax evasion on wage income is known to be very limited (Kleven, Knudsen, Kreiner, Pedersen, and Saez (2011)) we believe this data resource is accurate and precise. This registry has been in place since January 2008 and has previously been used for research purposes by Kreiner, Leth-Petersen, and Skov (2016) among others. We complement the registry with information about the receipt of public transfer income and with information about household composition and various background characteristics. Registry data are available for the entire population, and for the purposes of this study, the administrative data have been merged with all women to whom we have sent out invitations to participate. In Table A1 in the Appendix we display summary statistics for participants and non-participants. The Table shows that participants are slightly older, slightly more likely to have a partner and to have kids, be Danish citizens, have slightly higher income and have slightly longer educations. While these differences are often statistically significant, in most cases the quantitative differences are modest.

3.3 Results: Estimating MTEs Using Subjective Expectations Data

3.3.1 Subjective Expectations Data

Figure 2: $\hat{G}^i$, The Reported Probability of Being Treated (Having a Child in the Next 4 Years)



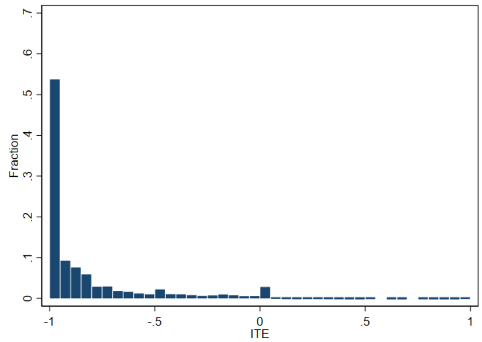
Notes: This figure plots a histogram of survey responses of the subjective probability of having a child over a 4-year horizon.

Key for estimating marginal treatment effects is the probability of having a child. In Figure 2 we present a histogram of the distribution of expected fertility within the four years following the survey. Importantly for estimating $MTEs$, child birth expectations are distributed across the entire unit interval.[8]

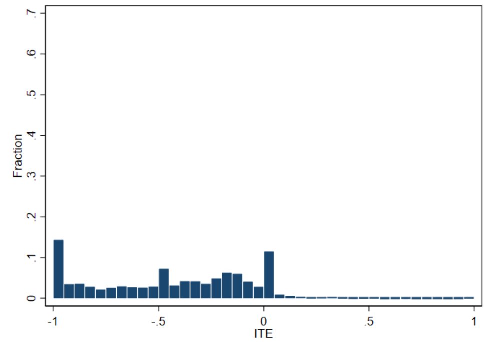
The other key subjective probability data is the subjective ex ante ITE, that is, the difference in expected labor supply conditional on having and not having a child. We present these $\widehat{seaITEs}$ in Figure 3. Figure 3a shows that most women typically work at baseline and most do not work 3 months after childbirth, i.e., the ITE is highly concentrated around -1. There is some ITE heterogeneity at month 3, but it is pretty limited, which means there is some limited scope for heterogeneity in 3-month MTEs across the population. Figure 3b shows that there is substantial ITE heterogeneity at month 9, which means there is scope for substantial heterogeneity in 9-month MTEs. Whether there will actually be heterogeneity

[8] The histogram appears to indicate that roughly half of the women in the sample report extreme probabilities, i.e., 0 or 100. This is, however, due to the coarseness of the bins. In fact, 17 percent of the women in the sample indicate zero percent probability of child birth and 9 percent of the women indicate 100 percent probability of child birth.

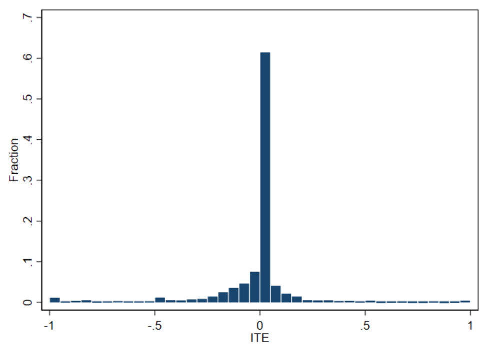
Figure 3: $\widehat{seaITE}$ Density



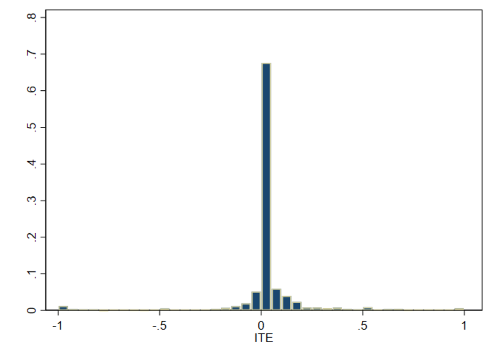
(a) ITE Density 3 Months after Birth



(b) ITE Density 9 Months after Birth



(c) ITE Density 18 Months after Birth



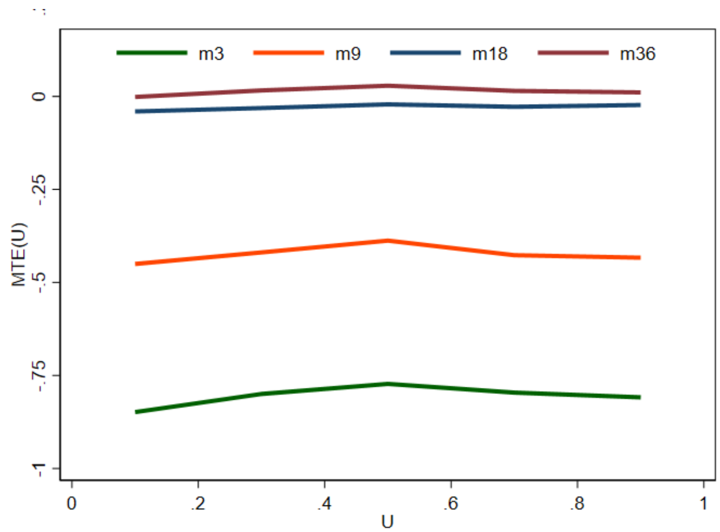
(d) ITE Density 36 Months after Birth

Notes: This figure plots a histogram of survey responses of the subjective ex ante individual treatment effect, computed as $\widehat{seaITE}^i_s = \widehat{\mathbb{E}}^i\left[Y^i_{1,s}\right] - \widehat{\mathbb{E}}^i\left[Y^i_{0,s}\right]$.

in 9-month MTEs depends on how the latent propensity to select into treatment correlates with the ITEs. Finally, Figures 3c and 3d show that there is very limited ITE heterogeneity at longer horizons. Almost all women report that they are equally likely to be working either in a typical month if they do not have a child or 18 or 36 months after having a child. This means there is limited scope for heterogeneity in long-run MTEs.

In Appendix Table D we regress $ITEs$ for months 3, 9, 8, and 36 after child birth on observable characteristics measured in the administrative registries in 2018, i.e., the year before the survey. There is systematic variation in who expects to return to work within 3 and 9 months after child birth. For example, women who already have children are more likely to return later than women who do not already have children and returning to work is postponed for more educated women at all horizons. If reported beliefs, $\widehat{\mathbb{E}}^i\left[Y^i_{1,s}\right]$, and $\widehat{\mathbb{E}}^i\left[Y^i_{0,s}\right]$, were mere noise, i.e., uninformative measures of the true subjective beliefs, $\mathbb{E}^i\left[Y^i_{1,s}\right]$, and $\mathbb{E}^i\left[Y^i_{0,s}\right]$, we would not expect to see them covary systematically with relevant covariates measured in

Figure 4: Subjective Ex Ante Marginal Treatment Effects ($\widehat{seaMTE_s}$)



Notes: Figure 4 shows marginal treatment effects calculated for various horizons after child birth. We order individuals by their $\hat{U}^i$ and calculate the average ITE within five equally sized buckets of $\hat{U}^i$.

administrative registries. We now turn to examine whether $ITEs$ vary systematically with the distaste for treatment, $\hat{U}^i$.

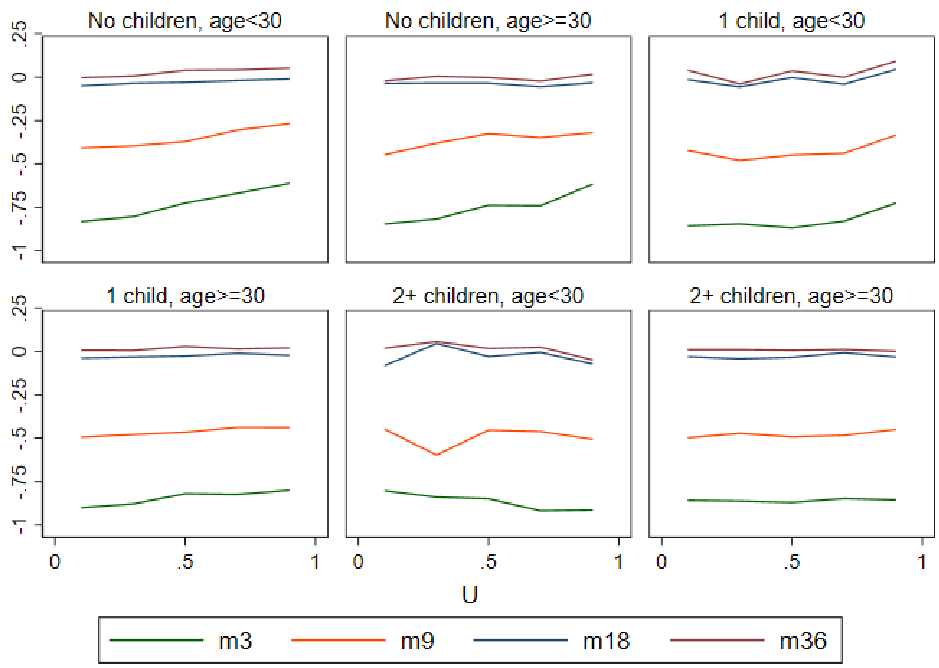
3.3.2 Estimating MTEs

Proposition 2 provides guidance for an estimator of MTEs that relies on the subjective probability data presented in Figures 2 and 3. In practice, we first select some X variables (e.g., age and number of children). Then we create a partition over X. For example, (age $\leq$ 30, no child); (age $>$ 30, no child); (age $\leq$ 30, 1 child); et cetera. Within each subset of the X partition, we order individuals according to $\hat{U}^i$ and then partition over $\hat{U}^i$, e.g., quartiles. Then, for each subset of the $\hat{U} \times X$ partition, we compute $\widehat{MTE}(u, x)$ as the average $\widehat{seaITE}$ for members of that subset. Once we compute the $\widehat{MTE}$s, we can compute $\widehat{ATE}$s, $\widehat{LATE}$s, $\widehat{ATT}$s, and $\widehat{ATU}$s by appropriately weighing the $\widehat{MTE}$ functions by $\hat{G}^i$, as outlined in Heckman and Vytlacil (2005). For example, the ATE is just the unconditional average ITE; the ATT is the weighted-average of the ITEs, with more weight on people more likely to be treated: $\left(\omega^i = \frac{\hat{G}^i}{\text{mean}(\hat{G}^i)}\right)$.

In Figure 4 we plot the estimated $\widehat{seaMTE}$, at this point without partitioning X into subsets.[9] The figure displays the estimated $\widehat{seaMTE_s}$ for the different months around child birth featured in the survey. We find that respondents anticipate different effects at differ-

[9]In Figure 4 we have binned $\hat{U}$ into five equally spaced groups. The results are robust to binning into ten groups. To avoid clutter we have omitted standard errors. In all cases, standard errors are very small.

Figure 5: Subjective Marginal Treatment Effects Conditional on Age and Number of Children



Notes: Figure 5 shows marginal treatment effects calculated for various horizons after child birth. We rank $\hat{U}^i$ within elements of the partition of the Age and Number of Children X variables and subsequently calculated the average $\widehat{ITE}$ within five equally sized buckets of $\hat{U}^i$ for each subset of the partition.

ent horizons, with respondents expecting about an 80-90 percentage point reduction in the probability of working 3 months after giving birth to a child, a 50 percentage point reduction after 9 months, and very modest effects on labor supply after 18 months and 36 months. The $\widehat{seaMTE}_s$ curves are quite flat, with some slight non-zero slope for months 3 and 9. Finding a pattern with limited selection on unobserved factors might not be unreasonable in a small and homogeneous society such as Denmark where a high level of income redistribution, large subsidies for publicly provided high-quality child care, and universal job protection laws work toward decreasing heterogeneity in opportunity costs.

It is possible that there may be $seaMTE$ heterogeneity according to observable characteristics X, and that computing $\widehat{seaMTE}$s conditional on $\hat{U}$, but not also conditional on subsets of X may be masking this heterogeneity. Thus, we explore the $\widehat{seaMTE}(u, x)$ function for a few X variables that have been demonstrated in the literature to be associated with working after childbirth. The specific Xs chosen are interesting economically, but also serve the purpose of exposition to demonstrate how to estimate $seaMTE(u, x)$ nonparametrically.

As Figure 5 clearly shows, the main source of selection on unobservables is at the 3-month horizon for women with no other children. The positive slope of $\widehat{seaMTE}(u, x)$ at 3 months for women with no children means that among these women, those who are less likely to have a child are more likely to work 3 months after having a child. Otherwise, $\widehat{seaMTE}$s are remarkably constant across older and younger women, across women with 1, 2, or 3 other children, and across all women at longer horizons.

Overall, we find little evidence of wide-spread MTE heterogeneity, especially at longer horizons. This means that $LATE \approx ATT \approx ATE$ and that existing treatment effect estimates can be extrapolated across and beyond the regions of the population where they are identified with less concern.

3.4 Investigating the Validity of Assumptions: Comparing Survey-based Results with Historical Data

In this section we compare fertility and labor supply beliefs with historical data from administrative registries. We show that the average treatment effect on the treated (ATT) estimated using the $\widehat{seaMTE}$ equals that estimated from a traditional event-study approach. We view this as supportive evidence of the veracity of the assumptions required for the method's validity and as evidence that, in this context, $\widehat{seaMTE} = seaMTE = MTE$.

3.4.1 Subjective Beliefs and Realizations of Child Birth have Similar Conditional Correlations with Key Demographic and Economic Variables

The average stated expected probability of having a child in the survey is 40 percent. In order to compare this number to measured historical fertility we consider all women aged 20-40 in 2012 and follow them four to six years forward. For this group, the average fertility four years after is 29 percent, and it is 35 and 41 percent five and six years after. We take this to reflect that the level of stated expected fertility aligns reasonably well with the level of actual fertility realized historically.

In Table 1 we regress fertility on demographic and economic characteristics. In column (1) the dependent variable is realized fertility within 2018 for all women aged 20-40 in 2012 and characteristics are measured in 2012. In column (2) the dependent variable is stated expected fertility and characteristics are measured in the registry data in 2018. Considering first the results for the survey sample in column (2), expected fertility is negatively correlated with the number of children. Fertility is positively correlated with the level of education and is highest among women aged 26-30. Having a partner is positively correlated with expected fertility. Interestingly, there appears not to be a strong correlation with the level of income given the other covariates. In column (1) we find much the same pattern, although in some cases the size of the estimated parameters differ and standard errors are smaller due to the

Table 1: **Regression of Historical and Expected Fertility on Covariates**

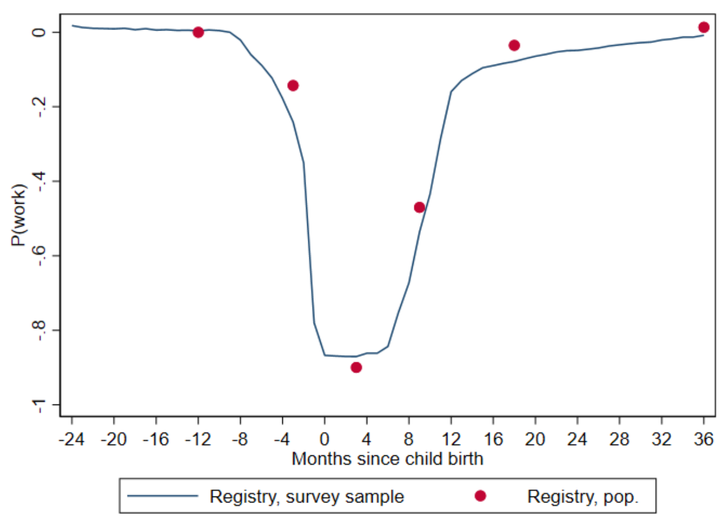
	(1) historical prob(child)	(2) subjective prob(child)
1 child	-0.117***	-0.002
	(0.014)	(0.013)
2 children	-0.459***	-0.316***
	(0.002)	(0.011)
3 children	-0.472***	-0.376***
	(0.002)	(0.013)
Educ, high school	-0.033***	-0.087***
	(0.002)	(0.018)
Educ, vocational	0.075***	-0.018
	(0.002)	(0.014)
Educ, middle	0.152***	0.039***
	(0.003)	(0.014)
Educ, college	0.169***	0.058***
	(0.003)	(0.014)
Age -25	-0.018***	-0.000
	(0.003)	(0.015)
Age 31-35	-0.050***	0.000
	(0.002)	(0.011)
Age 36-40	-0.325***	-0.193***
	(0.002)	(0.011)
Partner	0.334***	0.155***
	(0.002)	(0.010)
Income -200	-0.047***	0.004
	(0.002)	(0.011)
Income, 200-300	0.006**	0.026**
	(0.002)	(0.012)
Income, 400-500	-0.020***	-0.002
	(0.003)	(0.013)
Income, 500-	-0.045***	-0.030**
	(0.004)	(0.015)
Constant	0.439***	0.434***
	(0.003)	(0.017)
N	725,263	6,917

Notes: Column (1) presents a regression of the probability of having a child within 2018 for women aged 20-40 in 2012 according to the population registry. Regressors are measured in 2012. Column (2) presents a corresponding regression where the dependent variable is expected fertility in the baseline scenario from the survey. The regressors are from administrative registries and are measured in 2018. Robust standard errors in parentheses.* 10%, ** 5% *** 1% level of significance.

massive sample size. It is not possible using this data to directly verify that $\hat{G}^i$ preserves the rank of G^i as required by Assumption 3. However, the fact that the parameter estimates presented in Table 1, column (1) and (2), have the same signs and often also magnitude for each regressor is supportive evidence that fertility expectations are likely ranked similar to actual fertility chances.

3.4.2 ATTs from Subjective Probability Data are Quantitatively Close to ATTs from an Event Study using Realized Outcomes

Figure 6: Estimating Average Treatment Effect on the Treated (ATTs): Subjective Beliefs Method compared to Historical Event Study



Notes: The figure shows the percentage of women working at different horizons relative to child birth, where the fraction of women working in period -12 is set to zero. Estimates thus give the deviation in labor supply from the level at month -12. The red dots show estimates of the ATT based on survey responses. They are computed as the probability of working conditional on having a child averaged across respondents, weighting by the relative probability of having a child. The solid blue line shows event study ATT estimates, i.e., δ from equation (13), based on child births during the period 2010-2015 recorded in the administrative registries for the women who are in the survey sample.

Figure 6 displays with red dots the average treatment effects on the treated (ATTs) estimated using subjective probability data as detailed above. The red dots mark the expected percentage of women working at different horizons around child birth, where the fraction of women working in period -12 is set to zero such that estimates for other periods give the deviation in labor supply from the level at month -12. Expected labor supply starts decreasing three months before child birth, and most women expect to be out of work three months after the birth. The reason that expected labor supply does not drop to -1 is that not all women participating in the survey expect to be working 12 months before the expected child birth. Nine months after childbirth the average probability of working is about 50 percentage points below the level in month -12 and 18 months after birth average expected labor supply is only about three percentage points lower than at the outset. Three years after child birth labor supply has returned to the same level as before pregnancy.

In order to assess whether labor supply expectations around child birth are likely to reflect real behavior we compare survey responses to historical labor supply behavior. Specifically, we compare the ATTs estimated using our methodology with survey data to an estimate of ATTs using an event study approach with realized labor supply and fertility data. To do this we identify our survey respondents who have given birth to children in the period 2010-2015.[10] For these child births we follow the labor supply of the mother at a monthly frequency for the period covering 24 months before and up to 36 months after child birth. In order to quantify the labor supply response around child birth we employ an event study design by estimating the following regression:

$$y_{it} = \sum_{k \in K} D_{it}^{k} \delta_k + \gamma X_{it} + u_{it}. \tag{13}$$

y_{it} is a dummy variable taking the value one if individual i is working in month t. D_{it}^{k} is a dummy for the child being born k periods ago and δ_k measures the effect of the child birth k periods around birth. The month of conception is omitted, so $k \in K = \{-24, -23, \ldots, 36\} \setminus \{-9\}$. X is a vector of dummy variables controlling for the age of the mother and the timing of adjacent children that the mother has also given birth to. Women who give birth to more than one child during the period 2010–2015 enter with separate event time lines for each child.[11] Estimates of δ_k can be interpreted as the causal effect of having a child on labor supply in month k provided that the estimated effect prior to pregnancy, i.e., $k < -9$, is zero. Equation (13) is estimated on a sample of women who all have children but at different points in time, and δ_k can therefore be interpreted as an estimate of the average treatment effect on the treated (ATT). Thus, these event-study ATTs are directly comparable to the ATTs computed from estimated MTEs presented in Figure 6.

In Figure 6 we plot estimates of δ_k. The close alignment between the average survey

[10]See the Appendix C.2 for analysis using all women recorded in the *e-income* registry. Results are very similar, although expected labor supply is slightly higher than historical labor supply at longer horizons in the full population. The fact that the survey sample estimate is closer to expected labor supply could reflect that there is a modest level of self-selection into participating in the survey such that women who are more likely to return to work after having given birth are more likely to participate in the survey. Given the small deviations, we conclude, however, that such a pattern of self-selection is moderate.

[11]Appendix C contains further details about estimation. Estimation of equation 13 by OLS could be biased if treatment effects vary across treatment cohorts. In Appendix C.1 we present estimates based on the Sun and Abraham (2021) estimator that allows for such treatment effect heterogeneity. The results are practically identical to those presented in Figure 6.

responses and the responses estimated from administrative registry data document that the average treatment effect on the treated (ATT) estimated from the administrative data coincides with the ATT from subjective probability data in the (survey) population. We interpret the close alignment between the two ATT measures as strong supportive evidence that $\widehat{seaMTE} = seaMTE = MTE$ in this context.

To compute the survey-based ATTs we relied on Proposition 2' that required Assumptions 1, 2, 3, and 4 to hold. We find it unlikely that $\widehat{seaMTE} \neq seaMTE$, but $\widehat{seaMTE} = MTE$, so we think this alignment supports that individuals with similar observable characteristics use a similar model to forecast selection into treatment (Assumption 2) and that measurement issues (Assumptions 3–4) are not a concern, so $\widehat{seaMTE} = seaMTE$. Furthermore, we find it unlikely that the ATT computed using the $seaMTE$ and the ATT computed using historical data and an event study would be so close by coincidence. The ATT computed using subjective expectations is a weighted average of the $seaMTE$ and the event-study ATT is a weighted average of the MTE, so observing the same weighted average across approaches supports that the $seaMTE$ and MTE functions themselves are the same (although obviously this is not conclusive evidence).

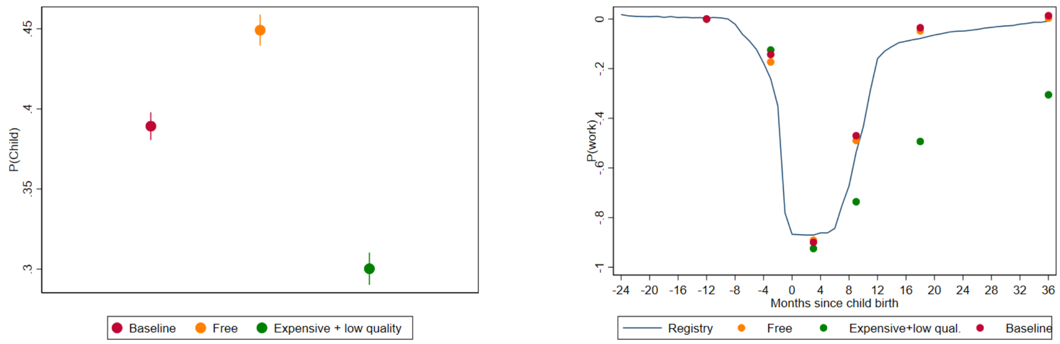
3.5 Counterfactual Analysis using $seaMTEs$

Of course, no method can possibly validate counterfactuals about policies that have never been observed without further assumptions. However, in the absence of a valid and relevant experiment, which is often the case when considering large-scale policy interventions, counterfactual analysis using $seaMTEs$ provides a data-driven methodology to forecast the effects of a hypothetical policy. This method can compliment predictions of pure theory by bringing data to the analysis and compliment structural models by requiring less reliance on extrapolation of model predictions beyond situations that informed the models' quantification.

To illustrate this, we ask about two hypothetical policies, one where we make child care free and another where we triple the price of child care while also lowering the quality of child care. The first policy change that we consider is a relatively small perturbation of the existing Danish child care system while the second hypothetical policy is a more drastic change that moves child care in Denmark closer to child care in the US. The policy scenarios are presented as follows:

- The first change we would like to ask you to consider is that as of today and in the foreseeable future, all municipal childcare is free. Assume that childcare quality is not affected by this cost change.

Figure 7: Counterfactual Policy Scenarios



(a) Childbirth Expectations

(b) ATT: Labor Supply

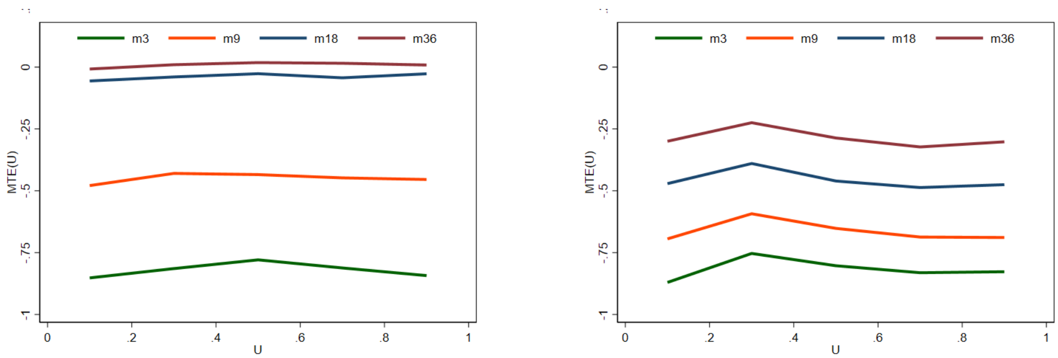
Notes: Panel a shows average expected fertility in the baseline scenario as well as in two hypothetical policy scenarios. In one scenario child care is made free and in the other scenario the prize of child care is tripled and the quality is reduced compared to the baseline. Panel b shows average estimated individual treatment effects for the baseline scenario and for the two hypothetical policy scenarios. The solid line shows event study estimates based based on child births recorded in the administrative registries for the women who are in the survey sample.

- The next change that we would like you to consider is that starting today and in perpetuity public child care is **low quality**. In particular, assume that child care staff was cut by 50% and that all employees are non-skilled. This would mean that no trained teachers would be employed, and that child care would focus on providing a place where the child can stay during work hours but not on child development.
  In addition, suppose that starting today and in perpetuity public child care was **more expensive**. In particular, suppose that the price of child care increased to be three times what it costs today. This would mean that the typical price of nursery care would increase to about 10,000 DKK per month for each child but that the quality is low as described above.

Within these alternative scenarios we ask questions about expected fertility and labor supply in exactly the same way as we did for the baseline case.

Figure 7 presents average expected fertility and labor supply for the baseline and for the two counterfactual scenarios. The free child care scenario is shown with yellow markers and the scenario with low quality and high price is shown with green markers. Making child care free increases expected fertility by about five percentage points and leaves labor supply conditional on having a child practically the same as in the baseline scenario. In contrast, lowering child care quality and at the same time tripling the price has a dramatic influence. Fertility drops by 10 percentage points compared to the baseline scenario, and labor supply is now persistently lower even after the job security/parental leave period runs out 12 months after birth. In fact, 18 months after birth labor supply conditional on having a child is 51

Figure 8: $\widehat{seaMTE}_s$ in Counterfactual Policy Scenarios



(a) Free Childcare (b) Expensive + Low Quality

Notes: Panel a shows marginal treatment effects calculated for various horizons after child birth for the hypothetical policy scenario where child care is made free. Panel b shows marginal treatment effects calculated for various horizons after child birth for the hypothetical policy scenario where the prize of child care is tripled and quality is lowered. For doing this we have ranked U within all combinations of the covariates that enter the regressions in Table 1 and subsequently calculated the average ITE within five equally sized buckets of U

percentage points lower than the pre-birth level, and after 36 months it is still 32 percentage points lower than the pre-birth level.

Table 2 presents regressions of fertility expectations and $\widehat{seaITEs}$ on covariates where fertility expectations and $\widehat{seaITEs}$ are collected for the counterfactual scenario where child care is expensive and of low quality.[12] In column (1) stated expected fertility in the expensive-low quality scenario is regressed on characteristics which are measured in the registry data in 2018. Qualitatively, results are comparable to the results for expected fertility in the baseline case, as can be seen in Table 1 column (2). In columns (2)-(5) the dependent variable is the $\widehat{seaITE}$ for month 3, 9, 18, and 36 after child birth. Unlike in the baseline case, there is considerable response heterogeneity. The regression results indicate that the decline in labor supply documented in Figure 7 is driven by women who already have children and those with higher education.

Figure 8 plots the $\widehat{seaMTE}$s for the two counterfactual scenarios, Figure 8a for the free child care scenario and Figure 8b for the scenario with high price and low quality. Comparing Figure 8a with the $\widehat{seaMTEs}$ from the baseline scenario in Figure 4 we see that the $MTEs$ are practically identical. Moreover, Figure 7a showed that making child care free increased fertility by five percentage points. Making child care free is thus an example

[12]Regressions for the baseline scenario corresponding to column (2)-(5) in Table 2 are shown in Appendix D. Note, there are fewer observations in the analysis of the counterfactual policy scenarios than in the baseline analysis. This is due to item-nonresponse. We have reproduced the baseline analysis using the subsample with valid responses to the questions pertaining to the counterfactual policy and find that the baseline scenario looks practically identical to the results presented in Figure 6.

Table 2: **Counterfactual Scenario with Expensive and Low Quality Child Care**

	(1)	(2)	(3)	(4)	(5)
	prob(child)	Month 3	Month 9	Month 18	Month 36
	b/se	b/se	b/se	b/se	b/se
1 child	-0.118***	-0.019	-0.051***	-0.086***	-0.078***
	(0.016)	(0.013)	(0.016)	(0.018)	(0.017)
2 children	-0.357***	-0.010	-0.056***	-0.077***	-0.080***
	(0.013)	(0.013)	(0.016)	(0.017)	(0.016)
3 children	-0.351***	0.003	-0.077***	-0.144***	-0.128***
	(0.016)	(0.019)	(0.022)	(0.025)	(0.024)
Educ, high school	-0.072***	-0.106***	-0.126***	-0.114***	-0.066**
	(0.022)	(0.025)	(0.027)	(0.028)	(0.026)
Educ, vocational	-0.037**	-0.113***	-0.065***	-0.003	0.011
	(0.017)	(0.021)	(0.023)	(0.023)	(0.021)
Educ, middle	0.021	-0.138***	-0.175***	-0.126***	-0.104***
	(0.018)	(0.020)	(0.022)	(0.024)	(0.022)
Educ, college	0.061***	-0.138***	-0.155***	-0.112***	-0.053**
	(0.018)	(0.020)	(0.022)	(0.023)	(0.021)
Age -25	-0.050***	0.013	0.030	0.036	0.058***
	(0.019)	(0.019)	(0.021)	(0.023)	(0.021)
Age 31-35	-0.015	0.047***	0.049***	0.044***	0.046***
	(0.013)	(0.012)	(0.015)	(0.017)	(0.016)
Age 36-40	-0.147***	0.035***	0.028*	0.009	0.023
	(0.013)	(0.013)	(0.016)	(0.018)	(0.017)
Partner	0.123***	-0.013	-0.015	-0.074***	-0.030**
	(0.013)	(0.012)	(0.015)	(0.016)	(0.015)
Income -200	-0.018	0.133***	0.107***	0.068***	0.035**
	(0.013)	(0.014)	(0.017)	(0.018)	(0.017)
Income, 200-300	0.006	0.042***	0.011	-0.001	0.007
	(0.014)	(0.013)	(0.017)	(0.019)	(0.018)
Income, 400-500	0.000	0.004	-0.003	0.003	0.018
	(0.015)	(0.013)	(0.017)	(0.019)	(0.018)
Income, 500-	-0.013	-0.020	0.008	0.039	0.045**
	(0.018)	(0.015)	(0.021)	(0.024)	(0.022)
Constant	0.412***	-0.762***	-0.567***	-0.319***	-0.220***
	(0.021)	(0.022)	(0.025)	(0.027)	(0.025)
N	5,098	4,970	4,963	4,956	4,968

Notes: Column (1) presents a corresponding regression where the dependent variable is expected fertility in the counterfactual scenario where child care is expensive and low quality. The regressors are from administrative registries and are measured in 2018. Columns (2)-(5) contain regressions where the dependent variable is the ITE pertaining to 3, 9, 18, and 36 months after child birth. Robust standard errors in parentheses.* 10%, ** 5% *** 1% level of significance.

of an alternative policy scenario where $MTEs$ are the same as in the baseline and where the behavioral response takes place through an adjustment in the propensity to be treated, which in this case is to have a child. Making child care free may be considered a relatively small policy change compared to the baseline in which child care is already heavily subsidized with 2/3 of the costs being covered by the local government. Changing child care to be of low quality and at the same time tripling the costs is a more dramatic change. Considering Figure 8b, it is obvious that $MTEs$ for this policy regime are quite different and lower than the baseline $MTEs$ shown in Figure 4.

Our counterfactual analysis based on $seaMTEs$ connects to the Heckman and Vytlacil (2005) concept of policy relevant treatment effects ($PRTE$). $PRTEs$ are the average impact on the outcome of interest caused by a change from the baseline policy to some alternative policy. $PRTEs$ can be constructed from baseline $MTEs$ as long as baseline $MTEs$ are policy invariant. This is the case if the policy affects the subjective treatment probability, G^i, but not the ex ante state-contingent expected outcomes $(\mathbb{E}^i[Y_1^i], \mathbb{E}^i[Y_0^i])$. In terms of our application, this means that the response to some alternative policy must only impact the propensity to have a child but not the labor supply response to having a child. It is typically difficult to test the policy invariance of estimated baseline $MTEs$ because naturally occurring data most often do not document behavior in alternative policy environments. However, an advantage of custom surveys is that it is possible to ask about the probability of treatment as well as state-contingent outcomes in counterfactual policy scenarios. The expensive and low quality childcare counterfactual is an example of a counterfactual policy scenario for which the baseline $MTEs$ cannot be considered policy invariant, and hence a counterfactual policy whose effects cannot be assessed from $PRTEs$ calculated from the baseline estimates.

4 Conclusion

The problem of external validity, i.e., extrapolating the effects of treatments estimated in one environment or sample to another, is an important but challenging problem. This paper proposes using surveys that measure subjective expectations to address this challenge. Working in the Generalized Roy model framework, we detail a combination of data and assumptions that identify the latent propensity to select into treatment and the subjective ex ante $MTEs$ across the entire population. We also show how purpose-designed survey data can be joined with typically measured historical behavioral data to assess the validity of these assumptions and when subjective ex ante marginal treatment effects equal traditional (objective) MTEs. Finally, we demonstrate how to study the effects of hypothetical policies, and their incidence across the population, before those policies are implemented.

We use the method to study the expected effects of childbirth on female labor supply in Denmark using a purpose-designed survey and administrative data. In the survey we measure the subjective expected probability of being treated and subjective expected individual treatment effects. We show that ex-ante average treatment effects and treatment beliefs align with treatment effects and average treatment probabilities estimated from historical administrative registry data. We compute average treatment effects on the treated (ATTs) from our estimated $seaMTE$s and show that they are near equal to ATTs estimated from an event study. We interpret these findings as supportive evidence that the assumptions underlying our approach are not violated and that subjective and objective MTEs are equal in this context.

The results indicate that $MTEs$ are roughly constant across levels of fertility expectations. These findings suggest that, after conditioning on typical demographic and economic characteristics, there is little selection into having children based on an idiosyncratic latent propensity to work after having a child. Having a child reduces the likelihood of working after birth by about 50 percent 9 months later, around 3 percent 18 months later, and around 0 percent 36 months later. We also study the effects of hypothetical policy interventions. Compared to baseline, in a counterfactual world in which childcare is lower quality and three times more expensive, fertility drops by 10 percentage points and the probability of working 36 months after childbirth is 32 percent lower.

Estimating $seaMTEs$ is useful as a complement to existing methods, particularly in cases where the treatment or outcomes have not yet materialized or where it is difficult to obtain observational data with sufficient quasi-experimental variation. Furthermore, our method allows us to understand the effect of treatment not only on outcomes but also on selection into treatment, thus giving a more complete picture of the likely real-life consequences of implementing new policies.

References

Aakvik, A., J. J. Heckman, and E. J. Vytlacil (2005): "Estimating treatment effects for discrete outcomes when responses to treatment vary: an application to Norwegian vocational rehabilitation programs," *Journal of Econometrics*, 125(1-2), 15–51.

Alesina, A., S. Stantcheva, and E. Teso (2018): "Intergenerational Mobility and Preferences for Redistribution," *American Economic Review*, 108(2), 521–554.

Andersen, H. Y., and S. Leth-Petersen (2021): "Housing Wealth Effects and Mortgage Borrowing: How Home Value Shocks Drive Home Equity Extraction and Spending," *Journal of the European Economic Association*, 19(1), 403–440.

Andersen, S. H. (2018): "Paternity Leave and the Motherhood Penalty:New Causal Evidence," *Journal of Marriage and Family*, 80, 1125–1143.

Angrist, I. D., and I. Fernandez-Val (2013): "ExtrapoLATE-ing: External Validity and," in *Advances in Economics and Econometrics: Volume 3, Econometrics: Tenth World Congress*, vol. 51, p. 401. Cambridge University Press.

Arcidiacono, P., V. J. Hotz, and S. Kang (2012): "Modeling college major choices using elicited measures of expectations and counterfactuals," *Journal of Econometrics*, 166(1), 3–16, Annals Issue on "Identification and Decisions", in Honor of Chuck Manski's 60th Birthday.

Arcidiacono, P., V. J. Hotz, A. Maurel, and T. Romano (2020): "Ex Ante Returns and Occupational Choice," *Journal of Political Economy*, 128(12), 4475–4522.

Attanasio, O. P. (2009): "Expectations and Perceptions in Developing Countries: Their Measurement and Their Use," *American Economic Review*, 99(2), 87–92.

Attanasio, O. P., and K. M. Kaufmann (2014): "Education choices and returns to schooling: Mothers' and youths' subjective expectations and their role by gender," *Journal of Development Economics*, 109, 203–216.

——— (2017): "Education choices and returns on the labor and marriage markets: Evidence from data on subjective expectations," *Journal of Economic Behavior and Organization*, 140, 35–55.

Bailey, M., E. Davila, T. Kuchler, and J. Stroebel (2019): "House Price Beliefs and Mortgage Leverage Choice," *Review of Economic Studies*, 86(6), 2403–2452.

Bjorklund, A., and R. Moffitt (1987): "The estimation of wage gains and welfare gains in self-selection models," *The Review of Economics and Statistics*, pp. 42–49.

Brinch, C. N., M. Mogstad, and M. Wiswall (2017): "Beyond LATE with a discrete instrument," *Journal of Political Economy*, 125(4), 985–1039.

Carneiro, P., J. J. Heckman, and E. J. Vytlacil (2011): "Estimating marginal returns to education," *American Economic Review*, 101(6), 2754–81.

Carneiro, P., M. Lokshin, and N. Umapathi (2017): "Average and marginal returns to upper secondary schooling in Indonesia," *Journal of Applied Econometrics*, 32(1), 16–36.

Christelis, D., D. Georgarakos, T. Jappelli, and M. van Rooij (2020): "Consumption Uncertainty and Precautionary Saving," *The Review of Economics and Statistics*, 102(1), 148–161.

Coibion, O., Y. Gorodnichenko, and T. Ropele (2019): "Inflation Expectations and Firm Decisions: New Causal Evidence*," *The Quarterly Journal of Economics*, 135(1), 165–219.

Colarieti, R., P. Mei, and S. Stantcheva (2024): "The How and Why of Household Reactions to Income Shocks," *working paper*.

Delavande, A. (2008): "PILL, PATCH, OR SHOT? SUBJECTIVE EXPECTATIONS AND BIRTH CONTROL CHOICE," *International Economic Review*, 49(3), 999–1042.

——— (2014): "Probabilistic Expectations in Developing Countries," *Annual Review of Economics*, 6, 1–20.

Delavande, A., and H.-P. Kohler (2016): "HIV-related Expectations and Risky Sexual Behavior in Malawi," *Review of Economic Studies*, 83(1), 118–164.

Delavande, A., and S. Rohwedder (2011): "Individuals Uncertainty about their Future Social Security Benefits and Portfolio Choice," *Journal of Applied Econometrics*, 26, 498–519.

Delavande, A., and B. Zafar (2019): "University Choice: The Role of Expected Earnings, Non-pecuniary Outcomes and Financial Constraints," *Journal of Political Economy*, 127(5), 2343–2393.

Dominitz, J., and C. F. Manski (1996): "Eliciting Student Expectations of the Returns to Schooling," *The Journal of Human Resources*, 31(1), 1–26.

Eisenhauer, P., J. J. Heckman, and E. Vytlacil (2015): "The Generalized Roy Model and the Cost-Benefit Analysis of Social Programs," *Journal of Political Economy*, 123(2), 413–443.

Fuster, A., G. Kaplan, and B. Zafar (2021): "What Would You Do with $500? Spending Responses to Gains, Losses, News, and Loans," *The Review of Economic Studies*, 88(4), 1760–1795.

Giustinelli, P., and C. F. Manski (2018): "Survey measures of family decision processes for econometric analysis of schooling decisions," *Economic Inquiry*, 56(1), 81–99.

Giustinelli, P., and M. D. Shapiro (2024): "SeaTE: Subjective Ex Ante Treatment Effect of Health on Retirement," *American Economic Journal: Applied Economics*, 16(2), 278–317.

Heckman, J., J. L. Tobias, and E. Vytlacil (2003): "Simple estimators for treatment parameters in a latent-variable framework," *Review of Economics and Statistics*, 85(3), 748–755.

Heckman, J. J., and R. Robb (1985): "Alternative methods for evaluating the impact of interventions: An overview," *Journal of Econometrics*, 30(1-2), 239–267.

Heckman, J. J., and E. J. Vytlacil (1999): "Local instrumental variables and latent variable models for identifying and bounding treatment effects," *Proceedings of the national Academy of Sciences*, 96(8), 4730–4734.

——— (2001): "Local Instrumental Variables," in *Nonlinear Statistical Modeling: Proceedings of the Thirteenth International Symposium in Economic Theory and Econometrics. Essays in Honor of Takeshi Amemiya*, pp. 1–46. Springer.

——— (2005): "Structural equations, treatment effects, and econometric policy evaluation," *Econometrica*, 73(3), 669–738.

——— (2007a): "Econometric evaluation of social programs, part I: Causal models, structural models and econometric policy evaluation," *Handbook of econometrics*, 6, 4779–4874.

——— (2007b): "Econometric evaluation of social programs, part II: Using the marginal treatment effect to organize alternative econometric estimators to evaluate social programs, and to forecast their effects in new environments," *Handbook of econometrics*, 6, 4875–5143.

Hendren, N. (2017): "Knowledge of future job loss and implications for unemployment insurance," *American Economic Review*, 107(7), 1778–1823.

Hudomiet, P., M. D. Hurd, A. M. Parker, and S. Rohwedder (2021): "The effects of job characteristics on retirement," *Journal of Pension Economics and Finance*, 20(SI3), 357–373.

Hurd, M. (2009): "Subjective Probabilities in Household Surveys," *Annual Review of Economics*, 1(2), 543–562.

Jappelli, T., and L. Pistaferri (2014): "Fiscal Policy and MPC Heterogeneity," *American Economic Journal: Macroeconomics*, 6(4), 107–136.

Kaufmann, K. M. (2014): "Understanding the income gradient in college attendance in Mexico: The role of heterogeneity in expected returns," *Quantitative Economics*, 5(3), 583–630.

Kleven, H., C. Landais, and J. E. Sogaard (2019): "Children and Gender Inequality: Evidence from Denmark," *American Economic Journal: Applied Economics*, 11(4), 181–209.

Kleven, H. J., M. B. Knudsen, C. T. Kreiner, S. Pedersen, and E. Saez (2011): "Unwilling or unable to cheat? Evidence from a tax audit experiment in Denmark," *Econometrica*, 79(3), 651–692.

Kreiner, C. T., S. Leth-Petersen, and P. E. Skov (2016): "Tax Reforms and Intertemporal Shifting of Wage Income: Evidence from Danish Monthly Payroll Records," *American Economic Journal: Economic Policy*, 8(3), 233–57.

Kuchler, T., and B. Zafar (2019): "Personal Experiences and Expectations about Aggregate Outcomes," *The Journal of Finance*, 74(5), 2491–2542.

Kuziemko, I., J. Pan, J. Shen, and E. Washington (2018): "The Mommy Effect: Do Women Anticipate the Employment Effects of Motherhood?," *working paper*.

Lochner, L. (2007): "Individual Perceptions of the Criminal Justice System," *American Economic Review*, 97(1), 444–460.

Lundborg, P., E. Plug, and A. W. Rasmussen (2017): "Can women have children and a career? IV evidence from IVF treatments," *American Economic Review*, 107(6), 1611–37.

Manski, C. F. (1990): "Nonparametric Bounds on Treatment Effects," *The American Economic Review*, 80(2), 319–323.

——— (1997): "Monotone Treatment Response," *Econometrica*, pp. 1311–1334.

——— (2003): *Partial identification of probability distributions*. Springer Science & Business Media.

——— (2004): "Measuring Expectations," *Econometrica*, 72(5), 1329–1376.

Manski, C. F., and J. V. Pepper (2000): "Monotone Instrumental Variables: With an Application to the Returns to Schooling," *Econometrica*, 68, 997–1010.

Mogstad, M., A. Santos, and A. Torgovitsky (2018): "Using instrumental variables for inference about policy relevant treatment parameters," *Econometrica*, 86(5), 1589–1619.

Pistaferri, L. (2001): "Superior information, income shocks and the permanent income hypothesis," *The Review of Economics and Statistics*, 83(3), 465–476.

Roy, A. D. (1951): "Some thoughts on the distribution of earnings," *Oxford economic papers*, 3(2), 135–146.

Rubin, D. B. (1974): "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies," *Journal of Educational Psychology*, 66(5), 688–701.

Stinebrickner, R., and T. R. Stinebrickner (2013): "A Major in Science? Initial Beliefs and Final Outcomes for College Major and Dropout," *The Review of Economic Studies*, 81(1), 426–472.

Sun, L., and S. Abraham (2021): "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects," *Journal of Econometrics*, 225(2), 175–199.

van der Klaauw, W. (2012): "On the Use of Expectations Data in Estimating Structural Dynamic Choice Models," *Journal of Labor Economics*, 30(3), 521–554.

Vytlacil, E. (2002): "Independence, monotonicity, and latent index models: An equivalence result," *Econometrica*, 70(1), 331–341.

Wiswall, M., and B. Zafar (2015): "Determinants of College Major Choice: Identification using an Information Experiment," *The Review of Economic Studies*, 82(2), 791–824.

——— (2020): "Human Capital Investments and Expectations about Career and Family," *Journal of Political Economy*, 129(5), 1361–1424.

Zafar, B. (2011): "How Do College Students Form Expectations?," *Journal of Labor Economics*, 29(2), 301–348.

——— (2013): "College Major Choice and the Gender Gap," *Journal of Human Resources*, 48(3), 545–595.

Appendix

A Proofs

Recall, Proposition 1' implies Proposition 1, Proposition 2' implies Proposition 2, and Corollary 1' implies Corollary 1.

Proposition 1'. *If Assumptions 1, 2, and 3 hold, then* $\hat{U}^i = \tilde{U}^i = U^i$.

Recall, $U^i := F_{V|X^i}(V^i)$, $F_{V|X^i}(v) := P(V \leq v|X^i = x)$, $G^i := P^i(D^i = 1)$, $\tilde{U}^i := 1 - F_{G|X^i}(G^i)$, and $\hat{U}^i := 1 - F_{\hat{G}|X^i}(\hat{G}^i)$.

Assumption 2. $G^i = 1 - \tilde{K}_{\mu|X^i}(V^i)$, *for some CDF* $\tilde{K}_{\mu|X^i}$.

Assumption 3. *Suppose that* $\hat{G}^i$ *is reported such that the order of individuals equals their order in the true subjective probabilities. That is,* $G^i < G^j \iff \hat{G}^i < \hat{G}^j \;\; \forall\, i, j$.

Proof. Assumption 3 immediately yields $\hat{U}^i = \tilde{U}^i$. Assumption 2 states that G^i is a strict monotonic decreasing transformation of V^i. This implies that individuals ordered by their G^i must have the same position in the order as when they are ordered by their V^i. Thus, $\tilde{U}^i = U^i$.

Proposition 2'. *If Assumptions 1, 2, 3, and 4 hold, then* $\widehat{seaMTE}(x, u) = seaMTE(x, u)$.

Recall the definition of the subjective ex ante MTE:

$$
\begin{aligned}
seaMTE(u, x) &:= \mathbb{E}\left[\mathbb{E}^i\left[Y_1^i\right] - \mathbb{E}^i\left[Y_0^i\right] |\tilde{U}^i = u, X^i = x\right], \\
&= \frac{1}{|\tilde{N}(u, x)|} \int \int \left[h_1^i(y)y - h_0^i(y)y\right] \; dy \; \mathbb{I}(i \in \tilde{N}(u, x)) \; di.
\end{aligned}
$$

the definition of the measured subjective ex ante ITE and $seaMTE$:

$$
\begin{aligned}
\widehat{seaITE}^i &:= \int \left[\hat{h}_1^i(y)y - \hat{h}_0^i(y)y\right] \; dy, \\
\widehat{seaMTE}(x, u) &:= \mathbb{E}\left[\widehat{seaITE}^i|X^i = x, \hat{U}^i = u\right],
\end{aligned}
$$

Assumption 4. *For all* i *and for* $D \in \{0, 1\}$, *for some deterministic function* $C^i(y)$ *and random variables* $\eta_D^i(y)$,

$$
\begin{aligned}
&\hat{h}_D^i(y) - h_D^i(y) = C^i(y) + \eta_D^i(y), \\
&\mathbb{E}\left[\eta_D^i(y)\right] = 0 \; \forall \; y.
\end{aligned}
$$

Proof.

$$
\begin{aligned}
\widehat{seaITE}^i &= \int \left[\hat{h}_1^i(y)y - \hat{h}_0^i(y)y\right] \; dy \\
&= \int \left[\left(h_1^i(y) + C^i(y) + \eta_1^i(y)\right) - \left(h_0^i(y) + C^i(y) + \eta_0^i(y)\right)\right] y \; dy \quad \text{(by Assumption 4)} \\
&= \int \left[\left(h_1^i(y) + \eta_1^i(y)\right) - \left(h_0^i(y) + \eta_0^i(y)\right)\right] y \; dy.
\end{aligned}
$$

Substituting this expression for $\widehat{seaITE}^i$ into the definition of $\widehat{seaMTE}(x,u)$ yields

$$
\begin{aligned}
\widehat{seaMTE}(x,u) &= \mathbb{E}\left[\int \left[\left(h_1^i(y) + \eta_1^i(y)\right) - \left(h_0^i(y) + \eta_0^i(y)\right)\right] y \; dy | X^i = x, \hat{U}^i = u\right] \\
&= \mathbb{E}\left[\int \left[h_1^i(y) - h_0^i(y)\right] y \; dy | X^i = x, \hat{U}^i = u\right] \quad \text{(by Assumption 4: } \mathbb{E}\left[\eta_D^i(y) = 0\right]) \\
&= \mathbb{E}\left[\int \left[h_1^i(y) - h_0^i(y)\right] y \; dy | X^i = x, \tilde{U}^i = u\right] \quad \text{(by Proposition 1')} \\
&= \frac{1}{|\tilde{N}(u,x)|} \int \int \left[h_1^i(y)\; y - h_0^i(y)\; y\right] \; dy \; \mathbb{I}(i \in \tilde{N}(u,x)) \; di, \\
&\text{thus, } \widehat{seaMTE}(x,u) = seaMTE(u,x).
\end{aligned}
$$

If $\eta_D^i(y) = 0 \; \forall \; y$, then $\widehat{seaITE}^i = seaITE^i$ even if $C^i(y) \neq 0$. When $\eta_D^i(y) \neq 0$, then generically $\widehat{seaITE}^i \neq seaITE^i$ because $[\eta_1^i(y) - \eta_0^i(y)] \neq 0$. However, averaging the mean-zero errors across individuals still yields $\widehat{seaMTE}(x,u) = seaMTE(u,x)$.

Corollary 1'. *Let Assumptions 1, 3, and 4 hold. If individuals have rational expectations, then* $\widehat{seaMTE}(x,u) = MTE(x,u)$. *This is because rational expectations implies* $seaMTE(x,u) = MTE(x,u)$.

B Balance Test for Danish Survey

Table A1: **Summary statistics for participants and nonparticipants**

Variable	(1) Non-participants	(2) Participants	(3) Difference
Age	28.624	29.413	0.789
	(6.606)	(6.454)	(0.069)
Partner	0.522	0.607	0.085
	(0.500)	(0.488)	(0.005)
Number of Children	0.700	0.782	0.082
	(1.045)	(1.026)	(0.011)
Had baby in 2017	0.034	0.039	0.005
	(0.180)	(0.193)	(0.002)
Non-DK citizenship	0.190	0.094	-0.096
	(0.392)	(0.292)	(0.003)
Income (1000DKK)	175.536	227.750	52.214
	(16.738)	(193.482)	(2.048)
Liquid assets (1000DKK)	74.988	87.297	12.309
	(484.323)	(206.467)	(3.120)
House owner	0.239	0.339	0.100
	(0.426)	(0.473)	(0.005)
Student	0.306	0.310	0.004
	(0.461)	(0.462)	(0.005)
Educ basic	0.220	0.125	-0.095
	(0.415)	(0.331)	(0.004)
Educ high school	0.206	0.184	-0.022
	(0.404)	(0.388)	(0.004)
Educ vocational	0.214	0.208	-0.005
	(0.410)	(0.406)	(0.004)
Educ middle	0.122	0.190	0.068
	(0.327)	(0.392)	(0.004)
Educ College	0.148	0.231	0.083
	(0.355)	(0.421)	(0.004)
Observations	39,336	11,312	50,648

Notes: Characteristics are obtained from administrative registries and are measured in 2018. Standard deviations in parentheses in columns (1) and (2). Standard errors in parenthesis in column (3).

C Estimating Average Treatment Effect on the Treated ($ATTs$): Subjective Beliefs Method compared to Historical Event Study

In Section 3.4.2 we estimate the ATT using an event-study regression. I practice, we separately estimate ATTs for women giving birth to their first, second and third child ($j = 1, 2, 3$) during 2010m1-2015m12. In each of these estimations we control for the age of the mother and for the timing of adjacent children that the mother has given birth to.

Specifically, the ATT k periods from child birth for women giving birth to their child number j is :

$$ATT_k^j = \mathbb{E}\left[y_{i,k}^j\right] \tag{14}$$

In practice, we estimate ATT_k^j using the following regression:

$$y_{it}^j = \sum_{\substack{k=-24 \\ k\neq-9}}^{36} D_{it}^{j,k}\delta_k^j + \gamma^j X_{it} + u_{it}^j \tag{15}$$

where y_{it} is a dummy variable taking the value one if individual i is working in month t. D_{it}^k is a dummy taking the value one if i gave birth to a child k months ago, and X is a vector of covariates including dummies for the age of the mother and event dummies for adjacent children. Including dummies for the timing of adjacent children is required to be able to estimate the direct effect of the child under consideration. For example, women who get a second child tend have the second child around two years after the first child, and estimating the effect of the first child without controlling for the timing of the second child would confound the effect of the first and the second child.

To arrive at an overall ATT estimate, we average the estimates using frequency weights.

$$ATT_k = \sum_j Pr(j) \times ATT_{j,k}$$

where $Pr(j)$ is the sample share of mothers giving birth to their jth child.

C.1 ATT estimates based on the Sun and Abraham (2021) estimator

The estimator proposed by Sun and Abraham first estimates the treatment-cohort specific average treatment effect on the treated and then weight these together using frequency weights. In our context, a treatment cohort consists of women giving birth to a child in a particular year × month cell. We include women giving birth to children during the period 2010m1-2015m12, i.e., the data include 60 treatment-cohorts.

Specifically, the ATT pertaining to k periods from child birth for cohort e is:

$$CATT_{e,k} = \mathbb{E}\left[y_{i,e,k}\right] \tag{16}$$

In practice we estimate $CATT_{e,k}$ using the following regression:

$$y_{it} = \sum_{e=2010m1}^{2015m12} \sum_{\substack{k=-24 \\ k\neq -9}}^{36} D_{it}^{e,k}\delta_k^e + \gamma X_{it} + u_{it} \tag{17}$$

where y_{it}^e is a dummy variable taking the value one if individual i belonging to treatment cohort e is working in month t. $D_{it}^{e,k}$ is a treatment-cohort specific dummy taking the value 1 in period k after child birth, and X is a vector of covariates including dummies for the age of the mother and event dummies for adjacent children.

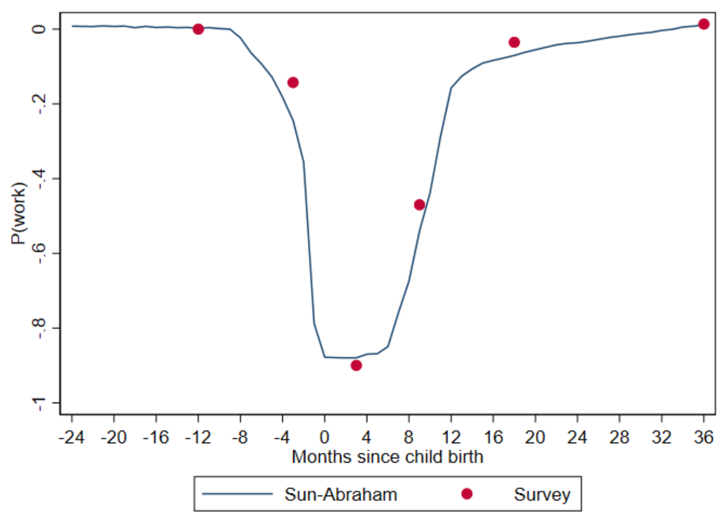
The average treatment effect k period from child birth, ATT_k, is calculated as the frequency weighted average of cohort specific effects, $CATT_{e,k}$

$$ATT_{e,k} = \sum_e Pr(e) \times CATT_{e,k}$$

where $Pr(e)$, is the sample shares of each cohort.

Implementing the Sun-Abraham estimator requires estimating many parameters. For each treatment cohort there are $K = 60$ event periods ($k = -24, ..., 36$) and $E = 60$ treatment cohorts ($e = 2010m1, ..., 2015m12$) and we thus need to estimate $(K-1)E = 3,540$ parameters. As we follow the procedure described in Appendix C, 3,540 parameters needs to be estimated for each child number ($j = 1, 2, 3$). In our data there are too few mothers who gave birth to a third child to be able to estimate $CATT_{e,k}$ for this group. We therefore include only observations for mothers giving birth to child number 1 and 2. Figure A1 shows the result. For all practical purposes, the figure is identical to Figure 6.

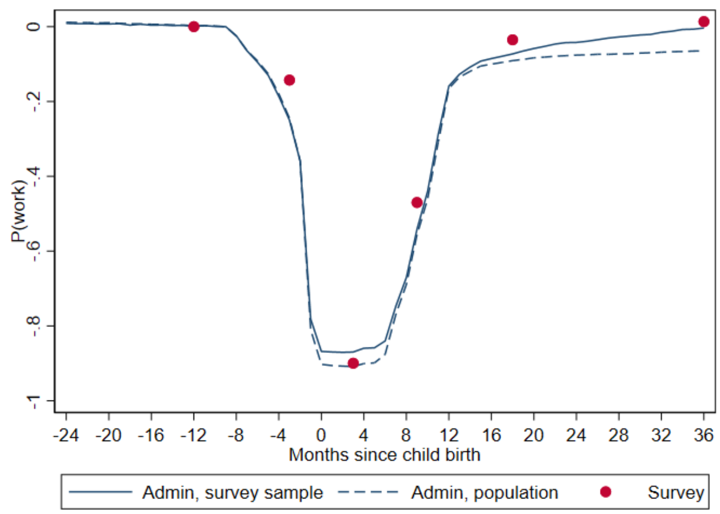
Figure A1: Estimating Average Treatment Effect on the Treated (ATTs): Subjective Beliefs Method compared to Historical Event Study Using The Sun-Abraham (2021) estimator



Notes: The figure shows the percentage of women working at different horizons relative to child birth, where the fraction of women working in period -12 is set to zero. Estimates thus give the deviation in labor supply from the level at month -12. The red dots show estimates of the ATT based on survey responses. They are computed as the probability of working conditional on having a child averaged across respondents, weighting by the relative probability of having a child. The solid blue line shows event study ATT estimates using the Sun and Abraham (2021). The sample includes women who are recorded in the administrative registries to have given birth during the period 2010-2015 and who participate in the survey.

C.2 Survey Sample versus Population

Figure A2: Estimating Average Treatment Effect on the Treated (ATTs): Subjective Beliefs Method compared to Historical Event Study



Notes: The Figure shows the percentage of women working at different horizons relative to child birth, where the fraction of women working in period -12 is set to zero. Estimates thus give the deviation in labor supply from the level at month -12. The red dots show estimates of the ATT based in survey responses. They are computed as the probability of working conditional on having a child averaged across respondents, weighting by the relative probability of having a child. The dashed blue line shows event study ATT estimates, i.e., δ from equation (13), based on registry data for the entire Danish population of women aged 20-40 who had children during the period 2010-2015. The solid line shows event study estimates based based on child births recorded in the administrative registries for the women who are in the survey sample.

D Correlation of ITEs and Covariates (Baseline Scenario)

Table A2: **How $ITEs$ correlate with Covariates in the Baseline Scenario**

	(1) Month 3	(2) Month 9	(3) Month 18	(4) Month 36
1 child	-0.089***	-0.091***	0.004	0.010
	(0.010)	(0.013)	(0.008)	(0.008)
2 children	-0.087***	-0.090***	0.007	0.010
	(0.010)	(0.013)	(0.008)	(0.007)
3 children	-0.068***	-0.105***	0.001	0.008
	(0.015)	(0.020)	(0.013)	(0.012)
Educ, high school	-0.078***	-0.057***	0.003	-0.000
	(0.018)	(0.019)	(0.013)	(0.013)
Educ, vocational	-0.101***	-0.055***	0.001	-0.033***
	(0.015)	(0.017)	(0.011)	(0.011)
Educ, middle	-0.156***	-0.152***	-0.022**	-0.042***
	(0.015)	(0.017)	(0.011)	(0.011)
Educ, college	-0.150***	-0.089***	-0.021*	-0.038***
	(0.015)	(0.016)	(0.011)	(0.011)
Age -25	-0.007	0.027*	-0.004	0.007
	(0.014)	(0.016)	(0.010)	(0.011)
Age 31-35	0.024**	0.020	-0.001	-0.001
	(0.010)	(0.012)	(0.008)	(0.007)
Age 36-40	0.048***	0.024*	0.000	0.005
	(0.010)	(0.013)	(0.008)	(0.008)
Partner	-0.030***	-0.015	-0.003	0.013*
	(0.010)	(0.011)	(0.007)	(0.007)
Income -200	0.149***	0.065***	0.006	0.039***
	(0.011)	(0.013)	(0.008)	(0.008)
Income, 200-300	0.033***	-0.004	0.002	0.014
	(0.010)	(0.014)	(0.009)	(0.009)
Income, 400-500	-0.017*	0.030**	0.012	0.004
	(0.009)	(0.015)	(0.008)	(0.008)
Income, 500-	-0.002	0.073***	0.014	0.000
	(0.013)	(0.018)	(0.010)	(0.009)
Constant	-0.694***	-0.329***	-0.023*	0.009
	(0.017)	(0.019)	(0.012)	(0.012)
N	6,688	6,645	6,669	6,638

Notes: The table shows regressions of $ITEs$ pertaining to 3, 9, 18, and 36 months after child birth on covariates. $ITEs$ are calculated from the survey responses, and the covariates are from administrative registries and are measured in 2018. Robust standard errors in parentheses.* 10%, ** 5% *** 1% level of significance.