The determinants of counterfactual identification in the binary choice model with endogenous regressors*

Telmo J. Pérez-Izquierdo†

November 12, 2022

Abstract

The Counterfactual Average Structural Function (CASF) is the Average Structural Function (ASF) averaged with respect to a counterfactual distribution for covariates. While the ASF is irregularly identified (not root-n estimable) for continuous regressors, the CASF may be regularly identified. This paper shows that, under a control function assumption, the CASF is non-parametrically identified as a weighted average. The weight is given by the likelihood ratio between the counterfactual density and the conditional density of the regressors given the control variables. Using this identifying moment condition, I obtain a necessary condition for regular identification of the CASF. The necessary condition depends on instrument strength, the degree of endogeneity, and the relevance of the regressors (i.e., how sensitive the outcome is to changes in regressors). Moreover, for normal DGPs, the necessary condition maps to a restriction on the structural parameters of the model. This provides further insights about what determines regular identification of the CASF. For instance, I find that if the first-stage R-squared is below a certain threshold, regular identification of the CASF is not possible.

Keywords: Non-parametric identification; Counterfactual analysis; Endogeneity; Control function.

JEL classification: C14; C21; C25.

*Research founded by Ministerio de Ciencia e Innovación, grant ECO2017-86675-P, MCI/AEI/FEDER/UE, grant PGC 2018-096732-B-100, and Comunidad de Madrid, grants EPUC3M11 (V PRICIT) and H2019/HUM-589.

†Universidad Carlos III de Madrid. Email address: teperezi@eco.ucm3.es.

1 Introduction

It is well established that the conditional expectation of the outcome given covariates does not recover relevant structural information for non-linear models with endogenous regressors (Blundell and Powell, 2003; Lin and Wooldridge, 2015). For those settings, Blundell and Powell (2003) introduced the Average Structural Function (ASF).[1] The ASF "would be the counterfactual conditional expectation of the outcome given regressors if the endogeneity of regressors were absent, that is, if the regressors could be manipulated independently of errors" (Blundell and Powell, 2003, p. 316). Knowledge of the ASF is thus sufficient to study the effect of certain policy interventions (Blundell and Powell, 2003) and to obtain partial effects (Lin and Wooldridge, 2015). However, the ASF is a partial mean and, as such, it is irregularly identified for continuous regressors (Newey, 1994b; Imbens and Newey, 2009). Hence, estimation of the ASF at a parametric root-n rate is generally not possible (Chamberlain, 1986).

This paper studies the Counterfactual ASF (CASF) in the binary choice model with continuous endogenous regressors. This parameter was first introduced in Blundell and Powell (2004). The CASF consists on averaging out the ASF with respect to a counterfactual distribution for the regressors. The further integration step may be enough to regularize the problem, so that the CASF may be regularly identified (that is, identified with a positive information matrix, see Khan and Tamer, 2010; Escanciano, 2021). A natural question to ask then is whether this regularization happens for any counterfactual distribution. That is, whether the CASF is regularly identified for any counterfactual distribution (for which it is identified). Intuitively, the "further away" the counterfactual distribution is from the observed distribution, the harder it will be to identify the CASF. This paper formalizes this idea by obtaining and analyzing a necessary condition for regular identification of the CASF.

In the binary choice model, which is the focus of the present paper, the CASF can be read as the (population) proportion of "successes" when the distribution of regressors is exogenously set to a counterfactual. It is thus a relevant parameter for counterfactual analysis. The CASF can be used to perform Oaxaca (1973)-Blinder (1973) decompositions (Ao et al., 2021). Moreover, one can study the effects of a yet-to-implement policy that changes the distribution of regressors, as in Stock (1989, 1991). For instance, the empirical application in this paper studies the effect of economic growth on the prevalence of civil conflict in Sub-Saharan Africa. For this problem, the CASF gives the probability of conflict for a counterfactual distribution of economic growth. Thus, it can be useful to analyze yet-to-implement policies such as aid (which may shift the distribution of growth) or market integration (which may decrease its variance). Moreover, it could also serve to decompose the observed differences in conflict prevalence between Sub-Saharan Africa and other regions, such as the Middle East.

This paper shows that the CASF is non-parametrically identified by a moment condi-

[1] The ASF is also referred to as Average Dose-Response Function (see, e.g., Hirano and Imbens, 2004).

tion. Under a standard control function assumption, the CASF can be written as a weighted average of the outcome variable. The weight is given by the likelihood ratio between the counterfactual density and the conditional density of the regressors given the control variables. Moreover, I find that square-integrability of the moment condition that identifies the CASF is necessary for regular identification of the parameter. The proof strategy followed here is applicable beyond the CASF. It circumvents the problem of generated regressors, common to control function approaches. The strategy yields a necessary non-parametric condition for regular identification. Therefore, as regular identification is in turn necessary for regular root-n estimation (Chamberlain, 1986), the ideas of this paper may be used to rule out the later in further settings.

In the case of the CASF, the necessary condition for regular identification depends on the tails of two terms: the square of the weight and the conditional variance of the outcome given regressors and the control variable. This paper contributes by non-parametrically analyzing these terms, hence finding which factors determine whether the CASF is regularly identified.

For the weight to be square-integrable, it helps that the numerator is small relative to the denominator. The counterfactual density is in the numerator of the weight. Thus, the necessary condition for regular identification will be harder to satisfy for counterfactuals distant from the observed data: the denominator in the weight will be close to zero at points where the the counterfactual density (in the numerator) is still positive.

On the other hand, for a fixed counterfactual distribution, there are Data Generating Processes (DGPs) that satisfy the necessary condition for regular identification and others that do not. Square-integrability of the weight requires a heavy tailed denominator, that is, the conditional density of the regressors given control variables should have fat tails. This depends on how strong instruments are. Thus, regular identification of the CASF depends on instrument strength. Furthermore, for the binary choice model, the conditional variance of the outcome depends only on the “success” probability. This is influenced by two factors: relevance of regressors (how sensitive the outcome is to changes in regressors) and the degree of endogeneity.

To further illustrate how regular identification of the CASF depends on its determinants, I study how the non-parametric necessary condition behaves in normal DGPs. In this case, I am able to relate the necessary condition for regular identification with a condition on the structural parameters of the model. I find a bound on how large the variance of the counterfactual can be, relative to the observed one, for the CASF to be regularly identified. This confirms the difficulties arising if the counterfactual distribution is distant from the data. Moreover, the exercise provides a way to informally check whether the CASF is irregularly identified. For instance, I find that regular identification of the CASF is not possible if the first-stage R^2 is bellow a certain threshold. The threshold depends solely on the degree of endogeneity, with less endogenous regressors leading to a more stringent condition on instrument strength.

The paper is organized as follows. I begin with an overview of the related literature. Next, I introduce the binary choice model and define the ASF and the CASF. Section 4 provides the non-parametric moment condition that identifies the CASF. Regular identification of the CASF and its non-parametric determinants are studied in Section 5. In Section 6, I analyze the necessary condition for normal DGPs. Section 7 presents the empirical application on the effect of economic growth on conflict. Section 8 concludes.

2 Literature review

There is a large literature relating counterfactual analysis with the study of policy effects or decomposition analysis. The work by Stock (1989, 1991) focuses in non-parametric estimation of the effect that a shift in the distribution of covariates (i.e., a counterfactual distribution) has on the mean of the outcome variable. Regarding decomposition analysis, DiNardo et al. (1996) propose a reweighting scheme to estimate counterfactual densities in order to decompose the change in the US distribution of wages (for a survey, see Fortin et al., 2011).

Rothe (2010) and Chernozhukov et al. (2013) propose methods to estimate the entire distribution of the outcome when the distribution of exogenous covariates X_i is set to a counterfactual distribution P^{x*}. These authors assume a bounded support for the regressors and that the density of the regressors is bounded away from zero in their support. This rules out irregular identification. Instead, I study the possibly unbounded support case, where regular identification of the CASF is determined by the relative tail behavior of the factual and counterfactual distributions of the regressors.

The problem of irregular identification has recently been of concern in policy evaluation. For instance, Carneiro et al. (2010) propose to focus on a marginal version of the Policy Relevant Treatment Effect (Heckman and Vytlacil, 2005) as identification of the former is regular under certain conditions. Focusing on the CASF instead of the ASF follows a similar logic.

Moreover, Khan and Tamer (2010) show that the binary choice under Lewbel (1997)'s assumptions and the average treatment effect (Rubin, 1974; Rosenbaum and Rubin, 1983) are irregularly identified. In this case, they argue, irregularity comes from small denominators in the moment function. The two models studied by Khan and Tamer (2010) are identified by a weighted moment condition, where the weight is not square-integrable. This is the case of the CASF, which I show is non-parametrically identified by a weighted moment condition. Thus, the present paper complements Khan and Tamer (2010)'s results by finding a necessary condition for regular identification for another model and parameter.

Additionally, this paper contributes by analyzing which characteristics of the model influence regular identification of the CASF. Related to this is Imbens and Newey (2009), who link the estimation rate of the ASF to the decay rate of the joint distribution of the regressors and the control variable. For the Gaussian case, they discuss that the decay rate is linked to the first-stage R^2 (instrument strength). The present paper advances in this direction in two ways. First, I analyze the determinants of regular identification of the CASF non-parametrically. I find that, on top of instrument strength, relevance and endogeneity of the regressor also determine regular identification of the CASF. Second, I explicitly find the shape that the necessary condition takes for normal DGPs. This allows to better understand the channels through which regular identification of the CASF is affected by its determinants.

3 The binary choice model with endogenous regressors

3.1 Model and control function assumption

An independent and identically distributed sample $(Y_i, T_i, Z_{1i}, Z_{2i})_{i=1}^n$ is observed. To introduce the model, define $X_i \equiv (T_i, Z_{1i})$ and $Z_i \equiv (Z_{1i}, Z_{2i})$ as the vectors containing the variables in the structural and reduced-from equations, respectively. Finally, let U_i denote the structural error and V_i the reduced-form error or control variable. The distribution of $(Y_i, T_i, Z_{1i}, Z_{2i}, U_i, V_i)$ is P_0. The model governing the relationship between these random vectors is

$$\begin{aligned} Y_i &= \mathbb{1}\left(\theta_0(X_i) + U_i \geq 0\right), \\ T_i &= \pi_0(Z_i) + V_i, \end{aligned} \tag{3.1}$$

where $\mathbb{1}(E)$ takes value one if E is true and it is zero otherwise. I impose that $\mathrm{E}_{P_0}[V_i|Z_i] = 0$, so that $\pi_0(Z_i) = \mathrm{E}_{P_0}[T_i|Z_i]$. It is not assumed that U_i and V_i are independent, thus T_i is potentially endogenous in the the structural equation in (3.1). Regressors are allowed to enter in an unknown fashion, as in Matzkin (1992). That is, no parametric structure is imposed neither on the link function (θ_0), the reduced-form (π_0), nor in the joint distribution of (U_i, V_i) (the model is thus coined as non-parametric, c.f. Blundell and Powell, 2003).

In Section 4, I present identification of the CASF as a two step procedure: I first identify the ASF, which then leads to identification of the CASF. To cover the first step, I introduce a control function assumption. I assume that regressors (X_i) are independent of the structural error (U_i) conditional on the control function (V_i):

$$X_i \perp U_i | V_i. \tag{3.2}$$

The following assumption gathers up the restrictions imposed on the non-parametric binary choice model:

Assumption 3.1. $P_0 \in \mathcal{P}$, the set of probability measures P for $(Y_i, T_i, Z_{1i}, Z_{2i}, U_i, V_i)$ such that

a. The relationship given by (3.1) is satisfied for the functions $\theta_0 : \mathbb{R}^{\dim(X_i)} \to \mathbb{R}$ and $\pi_0 : \mathbb{R}^{\dim(Z_i)} \to \mathbb{R}^{\dim(T_i)}$.

b. $\mathrm{E}_{P_0}[V_i|Z_i] = 0$.

c. Condition (3.2) is satisfied.

This assumption is the cornerstone to identify the ASF, which will lead to identification of the CASF. Note, however, that Assumption 3.1 does not jointly identify (G_0, θ_0), being $G_0(u) \equiv P_0(-U_i \le u)$ the distribution of (minus) the structural error. To achieve this, Matzkin (1992) imposes additional assumptions on the set of possible functions among which one wants to identify (G_0, θ_0). Moreover, Assumption 3.1 is also not enough to identify θ_0 in case the link function is assumed to be linear (i.e., a semiparametric index). I refer to Blundell and Powell (2004) and Rothe (2009) for a discussion of identification of the index under the control function assumption (for a survey, see Lewbel et al., 2012).

3.2 The ASF and the CASF

The ASF conveys relevant structural information when regressors are endogenous (Blundell and Powell, 2003, 2004). For instance, the derivative of the ASF recovers the regression coefficients in the linear model with endogenous regressors (see Lin and Wooldridge, 2015). In the binary choice model, the ASF measures how the probability of a "success" (i.e., $Y_i = 1$) changes with the exogenous movement of X_i.

To define the ASF, consider that the policy maker exogenously assigns $X_i = x$ to an individual. According to the model, the outcome of this individual will be $Y_i(x) \equiv \mathbb{1}(\theta_0(x) + U_i \ge 0)$. To obtain the ASF, unobserved heterogeneity is averaged out:

$$\mathrm{ASF}_0(x) \equiv \mathrm{E}_{P_0}[Y_i(x)] = P_0(-U_i \le \theta_0(x)) = G_0(\theta_0(x)).$$

Blundell and Powell (2003) show that the ASF is identified as a partial mean (see also Section 4). Therefore, it is irregularly identified for continuous regressors (Newey, 1994b). The parameter of interest here is the Counterfactual ASF (CASF), which averages the ASF with respect to a counterfactual distribution. The further integration step may be enough to regularize the problem, so that the CASF may be regularly identified. This paper gives a necessary condition for regular identification of the CASF (see Section 5).

Let P^{x*} be a (counterfactual) distribution for regressors X_i. The CASF is defined as

$$\gamma_0^* \equiv \int \mathrm{ASF}_0(x) dP^{x*}(x) = \int G_0(\theta_0(x)) dP^{x*}(x). \tag{3.3}$$

As discussed bellow, this parameter is relevant for a bundle of empirical questions. When the ASF is invariant in the appropriate sense, the CASF measures the average response probability (the proportion of "successes") when the distribution of X_i is P^{x*}.

Invariance of the ASF is the endogenous regressor parallel of the invariance of conditional distributions assumption (see, for instance, the assumptions in Hsu et al., 2020; Ao et al., 2021). In the present setup, this means that both the "return" to observable characteristics (θ_0) and the distribution of unobserved heterogeneity (G_0) are invariant. It may be of interest to study counterfactual parameters that account for variation in these two terms.[2] However, this cannot be done in the non-parametric model given by Assumption 3.1, as θ_0 and G_0 are not identified.

Nevertheless, the CASF is useful for counterfactual analysis. For instance, if the ASF is also valid for a group with covariates distributed as P^{x*}, the CASF can be used to perform decomposition analysis. Moreover, under certain assumptions, the counterfactual effects estimated by decomposition methods can be interpreted as treatment effects (Fortin et al., 2011; Chernozhukov et al., 2013). This framework can also be used to study external validity (Hsu et al., 2020).

Furthermore, one can study a yet-to-implement policy that changes the distribution of regressors to P^{x*}, as in Stock (1989). If the ASF is unaffected by the policy, then the CASF measures the effect of the policy. In Hsu et al. (2020)'s terminology, the counterfactual distribution P^{x*} can be independent of the observed distribution of X_i; or it may be the outcome of a deterministic transformation $\ell(X_i)$, as in Imbens and Newey (2009). In the latter case, $P^{x*} = P_0^x \circ \ell^{-1}$, being P_0^x the distribution of X_i under P_0. This includes the trivial case, where ℓ is the identity and $P^{x*} = P_0^x$.

4 Identification of the CASF

The CASF depends on the counterfactual distribution for which the researches decides to compute it. In the present section, I characterize for which counterfactual distributions it is possible to identify the CASF. The section concludes showing that the CASF is non-parametrically identified by a moment condition.

4.1 Support conditions

The CASF is identified for counterfactual distributions satisfying certain support conditions. To characterize the set of counterfactuals for which the CASF is identified, the first step is to analyze identification of the ASF (for a discussion, see Blundell and Powell, 2003). Then, to identify the CASF, one must require that the counterfactual distribution does not put positive mass in points where the ASF is not identified.

Key for identification of the ASF is the control function assumption: conditional independence of U_i and X_i given $V_i = v$. This means that $Y_i(x)$ and X_i are also conditionally independent. With this in mind, one gets

$$\begin{aligned}\mathrm{ASF}_0(x) \equiv \mathrm{E}_{P_0}\left[Y_i(x)\right] = \mathrm{E}_{P_0}\left[\mathrm{E}_{P_0}\left[Y_i(x)|V_i\right]\right] = \mathrm{E}_{P_0}\left[\mathrm{E}_{P_0}\left[Y_i(x)|X_i = x, V_i\right]\right] = \\ = \mathrm{E}_{P_0}\left[\mathrm{E}_{P_0}\left[Y_i|X_i = x, V_i\right]\right] = \int \mathrm{E}_{P_0}\left[Y_i|X_i = x, V_i = v\right] dP_0^v(v),\end{aligned} \tag{4.1}$$

where P_0^v is the marginal distribution of V_i under P_0. Note that the control variable $V_i = T_i - \pi_0(Z_i)$ is "observed" as $\pi_0(z)$ is identified. Therefore, the ASF is identified as a partial

[2] I thank an anonymous referee for pointing this out.

mean for certain values of $x \in \mathbb{R}^{\dim(X_i)}$ (Blundell and Powell, 2003).

For which values is the ASF identified? In the above equation, when integrating with respect to the distribution of V_i, one must be able to compute the conditional expectation of Y_i through the hyperplane $\{x\} \times \sup(V_i)$, being $\sup(V_i)$ the support of V_i. Therefore, identification of the ASF for a given $x \in \mathbb{R}^{\dim(X_i)}$ requires that $\{x\} \times \sup(V_i) \subset \sup(X_i, V_i)$. This has led some authors to require that $\sup(V_i|X_i = x) = \sup(V_i)$ for the ASF to be identified at x (see Blundell and Powell, 2003; Imbens and Newey, 2009).

Here, I get a condition for the conditional distribution of X_i given $V_i = v$. This condition is equivalent to the one discussed in Blundell and Powell (2003). The usefulness of working with the conditional distribution of X_i given $V_i = v$ is that this can be linked with instrument strength, as discussed below. If, for almost all v, $x \in \sup(X_i|V_i = v)$, then the hyperplane $\{x\} \times \sup(V_i)$ will be contained in the support of (X_i, V_i). Therefore, the ASF is identified at any point in the set

$$\mathcal{S} \equiv \left\{x \in \mathbb{R}^{\dim(X_i)} : x \in \sup(X_i|V_i = v) \text{ a.s. } P_0^v\right\}. \tag{4.2}$$

To obtain the CASF, the researcher averages the ASF with respect to a predetermined counterfactual distribution of the regressors (see equation (3.3)). This distribution must not assign a positive probability to points at which the ASF is not identified. Hence, when the ASF is non-parametrically identified, the CASF is only identified for counterfactual distributions whose support lies within the above set. This result is highlighted in the following proposition:[3]

Proposition 4.1. *Assume that the distribution P_0 of $(Y_i, X_i, Z_{2i}, U_i, V_i)$ satisfies Assumption 3.1. Then, the CASF is identified if and only if the counterfactual distribution P^{x*} satisfies*

$$\sup_{P^{x*}}(X_i) \subseteq \mathcal{S},$$

where $\sup_{P^{x*}}(X_i)$ *refers to the support of X_i under P^{x*} and $\mathcal{S}$ is defined in equation* (4.2).

This proposition already highlights the relevance of instrument strength in non-parametric identification of the CASF. The size of $\mathcal{S}$ depends on the conditional distribution of X_i given $V_i = v$. If instruments are weak, knowledge of V_i will provide a lot of information about T_i (recall that $T_i = \pi_0(Z_i) + V_i$). Therefore, the conditional support of X_i will be small, leading to a tinier set of counterfactuals where the CASF is identified.

An illustrative case is the one where the instruments Z_{2i} have no explanatory power. That is, $\pi_0(z_1, z_2) = \pi_0(z_1)$. In this case, $\mathcal{S} = \emptyset$ if V_i is non-degenerate, as I now show. If π_0 is not a function of the instruments:

$$\sup(X_i|V_i = v) \subseteq \{(t, z_1) \in \sup(X_i) : t - \pi_0(z_1) = v\}.$$

[3] The proof of the results of this paper are given, in order of appearance in the text, in Appendix A.

Take now $v_1, v_2 \in \sup(V_i)$ such that $v_1 \neq v_2$. If $\mathcal{S}$ is non-empty, there is an $x = (t, z_1)$ satisfying that $x \in \sup(X_i|V_i = v)$ for $v \in \{v_1, v_2\}$. This leads to $v_1 = v_2$, contradicting the initial statement. That is, if instruments are not related to the endogenous variables (or if instruments are not excluded from the structural equation), the CASF is unidentified for every counterfactual distribution. This drastic case illustrates the dangers of counterfactual analysis under weak instruments.

It is also worth noting that Proposition 4.1 rules out discrete endogenous regressors with finite support (cf. Blundell and Powell, 2003). In that case $\mathcal{S} = \emptyset$, so the CASF is unidentified for any counterfactual distribution. To illustrate, consider there is a single endogenous dummy regressor T_i as in Vytlacil and Yildiz (2007). Then, $\pi_0(z) \in [0, 1]$ and, since $V_i = T_i - \pi_0(Z_i)$, $\sup(V_i) \subseteq [-1, 1]$. Now, if $V_i = v > 0$ (underprediction), then it must be that $T_i = 1$, as π_0 cannot be negative. The opposite happens when $V_i < 0$ (overprediction): $T_i = 0$, as π_0 is lesser than 1. Only $V_i = 0$ leaves one uncertain, as one could have perfectly predicted either $T_i = 0$ or $T_i = 1$. Hence, for any $v \in \sup(V_i)$,

$$\sup(T_i|V_i = v) = \begin{cases} \{0\} & \text{if } v < 0, \\ \{0, 1\} & \text{if } v = 0, \\ \{1\} & \text{if } v > 0. \end{cases}$$

That is, $\sup(T_i|V_i = v_0) \cap \sup(T_i|V_i = v_1) = \emptyset$ for $v_0 < 0$ and $v_1 > 0$. Therefore, if V_i takes both positive and negative values, which is generally the case, $\mathcal{S}$ is empty and the CASF is not identified.

4.2 An identifying moment condition for the CASF

When the CASF is identified, it can be written as a moment. Following the definition of the parameter and the identification result in equation (4.1), the CASF can be expressed as

$$\gamma_0^* \equiv \int \mathrm{ASF}_0(x) dP^{x*}(x) = \iint \mathrm{E}_{P_0}\left[Y_i | X_i = x, V_i = v\right] dP_0^v(v) dP^{x*}(x). \tag{4.3}$$

The above integral is computed with respect to the product measure $P_0^v \times P^{x*}$. To obtain a moment condition that non-parametrically identifies the CASF, one must perform a change of measure. This can be done under the identifying assumption:

Assumption 4.1. The support of the counterfactual distribution satisfies $\sup_{P^{x*}}(X_i) \subseteq \mathcal{S}$.

This assumption states that the counterfactual distribution satisfies the condition in Proposition 4.1. It guarantees that the measure $P_0^v \times P^{x*}$ is absolutely continuous with respect to P_0^{xv}, the joint distribution of (X_i, V_i).

To express the moment condition in terms of a ratio of densities, I impose an additional assumption. I require that (X_i, V_i) has a density with respect to a product measure:

Assumption 4.2. The random vector (X_i, V_i) has density f_0^{xv} w.r.t. a σ-finite measure $\nu = \nu^x \times \nu^v$ on $\mathbb{R}^{\dim(X_i)} \times \mathbb{R}^{\dim(T_i)}$.

This assumption is far from being restrictive. Assumption 4.2 is satisfied by any random vector (X_i, V_i) with joint density with respect to products of the Lebesgue and counting measures. This is usually the case when continuous and categorical regressors are present in the model. Note also that, since $\mathcal{S} \subseteq \sup(X_i)$, the preceding two assumptions imply that P^{x*} has density w.r.t. ν^x. I denote this density by f^{x*}.

Under the above assumptions, the change of measures is allowed and the CASF can be written as a weighted average of the outcome variable Y_i:

Proposition 4.2. *Assume that the distribution P_0 of $(Y_i, X_i, Z_{2i}, U_i, V_i)$ satisfies Assumptions 3.1, 4.1, and 4.2. Then, the CASF is identified by*

$$\gamma_0^* = \mathrm{E}_{P_0}\left[Y_i \alpha_0(X_i, V_i)\right], \tag{4.4}$$

where the weight α_0 is given by

$$\alpha_0(x, v) \equiv \frac{f_0^v(v) f^{x*}(x)}{f_0^{xv}(x, v)} = \frac{f^{x*}(x)}{f_0^{x|v}(x, v)}, \tag{4.5}$$

being $f_0^{x|v}(x, v)$ and $f_0^v(v)$ the conditional density of X_i given $V_i = v$ and the marginal density of V_i under P_0, respectively.

Since $V_i = T_i - \pi_0(Z_i)$, equation (4.4) identifies the CASF as a moment condition: $\gamma_0^* = \mathrm{E}_{P_0}\left[Y_i \alpha_0(X_i, T_i - \pi_0(Z_i))\right]$. Under the control function assumption, identification of π_0 in the first-step equation leads to identifying the CASF (note that α_0 is identified once π_0 is). Hence, the CASF is identified irrespectively of whether θ_0 is identified. This means the CASF is identified even if the whole structural model is not.

The weight α_0 is the likelihood ratio between the counterfactual distribution of X_i and the conditional distribution of X_i given $V_i = v$. It highlights the relevance, regarding identification of the CASF, of the conditional distribution of X_i, instead of its marginal distribution. As I further discuss in the next section, this finding reveals the important role of instrument strength. Indeed, the presence of the conditional distribution comes from the endogeneity of T_i. To see this, consider the case where T_i is exogenous. That is, we may assume that $X_i \perp U_i$. Then, the weight in the moment condition would be f^{x*}/f_0^x, which depends solely on the factual and counterfactual distributions of X_i.

5 Regular identification of the CASF

The ASF is identified as a partial mean. Therefore, the ASF at a certain point $x \in \mathcal{S}$ is irregularly identified (i.e., it is identified but its Fisher information is zero). This, for instance, means that the convergence rate of any estimator will be slower than root-n (Chamberlain, 1986). On the other hand, the CASF is a full mean: an average over all components. As such, it may be regularly identified and have a finite information bound (for some results, see Newey, 1994b). This section finds a necessary condition for the CASF to be regularly identified. This condition imposes a restriction on the counterfactual distribution for which the CASF is computed. Moreover, the condition depends on various features of the observed data, such as instrument strength.

To obtain the necessary condition, I rely on two results. The first is the identifying moment condition given in Proposition 4.1. The second component of the strategy is to study a specific submodel ($\mathcal{P}'$), instead of the whole model $\mathcal{P}$ given by Assumption 3.1 (c.f. Komunjer and Vuong, 2010, who also focus on a submodel, which is parametric in their case). Indeed, estimation of the CASF in any submodel is easier than in the whole model. In the former, the researcher has more information at her disposal. Thus, the information bound for the CASF in the submodel will be no smaller than the one computed for the entire model. The following lemma formalizes this intuitive fact:

Lemma 5.1. *Consider two models, $\mathcal{P}$ and $\mathcal{P}'$, where $\mathcal{P}' \subseteq \mathcal{P}$. Then, the information bound for estimating γ_0^* in $\mathcal{P}$ is no smaller than the one for estimating γ_0^* in $\mathcal{P}'$.*

To estimate the CASF using the moment condition, the researcher must fit a reduced-form for T_i and estimate the weight α_0 prior to computation of the weighted average. The semiparametric information bound for this problem, where the control variables V_i must be generated, is not trivial. Instead, in the submodel where the control variables are observed and the distribution of (X_i, V_i) is known to the researcher, estimation of the CASF reduces to computing a weighted average (where the weight α_0 is known). A necessary condition to find a regular estimator of the expectation of the random variable $Y_i\alpha_0(X_i, V_i)$ is that it has finite variance.

I therefore obtain the semiparametric efficiency bound for the CASF in the submodel where V_i is observed and the weight α_0 is known. By the above lemma, the efficiency bound in the whole model is greater or equal to the one in the submodel. Therefore, the efficiency bound in the submodel must be finite for the CASF to be regularly identified. This strategy leads to a clear and easy to interpret necessary condition for regular identification of the CASF in the whole model. The following theorem states the result:

Theorem 5.1. *Assume that the distribution P_0 of $(Y_i, X_i, Z_{2i}, U_i, V_i)$ satisfies Assumptions 3.1, 4.1, and 4.2. Let α_0 be given by equation (4.5) and $\sigma_Y^2(x, v) \equiv \mathrm{Var}(Y_i|X_i = x, V_i = v)$ denote the conditional variance of Y_i given (X_i, V_i) under P_0. If the counterfactual distribution does not satisfy*

$$\mathrm{E}_{P_0}\left[\alpha_0(X_i, V_i)^2\sigma_Y^2(X_i, V_i)\right] < \infty, \tag{5.1}$$

then, the information bound for the CASF is infinite and it is irregularly identified.

Note that no additional assumptions are imposed in the whole model $\mathcal{P}$, so the condition in equation (5.1) is necessary. I conjecture that the above condition is, however, not sufficient (a formal proof of this results is out of the scope of this paper). The above condition disregards the effect of the first-stage regression on estimation of the CASF. This increases, in general, the asymptotic variance of any estimator of the CASF (an exception is adaptive estimation, see Bickel et al., 1998). Therefore, it may be that the condition in equation (5.1) is satisfied, but that the additional noise from the first-stage regression makes regular estimation of the CASF impossible.

Regarding estimation of the CASF, the identifying moment condition in equation (4.4) may be useful. A pathwise derivative calculation, as in Newey (1994a), can be used to find the asymptotic variance of the resulting estimator. However, the effect of generated regressors must be taken into consideration. This requires to extend the results in Hahn and Ridder (2013, 2019) to densities. Moreover, a trimming sequence may be necessary if the CASF is irregularly identified (see Khan and Tamer, 2010). Hence, efficient estimation of the CASF in the non-parametric model is left for future research.

5.1 Non-parametric determinants of counterfactual identification

Whether the necessary condition for identification of the CASF is satisfied depends on the counterfactual distribution (P^{x*}) and the DGP (P_0). For a given counterfactual distribution, the researcher is uncertain of whether the CASF is regularly or irregularly identified (compare, for instance, with Khan and Tamer, 2010, where their parameters are known to be irregularly identified). Indeed, the analysis of normal DGPs in Section 6 shows that, for a given counterfactual, there are DGPs in $\mathcal{P}$ for which the condition is satisfied, and other for which it is not.

In this section, I analyze which factors determine whether the necessary condition is satisfied. That is, which characteristics of the DGP determine whether a given counterfactual may be regularly identified. I use "counterfactual identification" as a shortcut for regular identification of the CASF for the given counterfactual. I separately study the two terms that appear in the necessary condition: the weight α_0 and the conditional variance of the outcome σ_Y^2.

Instrument strength: The weight α_0 depends solely in the first-stage regression and the counterfactual density. For a fixed counterfactual distribution f^{x*}, α_0 depends on the joint distribution of regressors and instruments: (X_i, Z_{2i}). Thus, instrument strength is the relevant measure to understand the behavior of the weight. Key to identifying the CASF is that, under the control function assumption,

$$\begin{aligned}\mathrm{E}_{P_0}\left[Y_i|X_i=x,V_i=v\right] = \mathrm{E}_{P_0}\left[Y_i(x)|X_i=x,V_i=v\right] = \mathrm{E}_{P_0}\left[Y_i(x)|V_i=v\right] = \\ = P_0(-U_i \leq \theta_0(x)|V_i=v) = F_0(\theta_0(x),v),\end{aligned}$$

where $F_0(u,v) \equiv P_0(-U_i \leq u|V_i=v)$ is the conditional distribution of (minus) the structural error given the control variable. That is, one can identify $F_0(\theta_0(x),v)$ as X_i moves exogenously once one has conditioned on $V_i=v$. However, if $V_i=v$ is fixed, variation of X_i must come from varying the instrument, Z_{2i}.

Theorem 5.1 formalizes the above intuition. Regular identification of the CASF is related to square-integrability of the weight α_0. Hence, it imposes a restriction on the tail behavior of the counterfactual distribution relative to the tails of the conditional distribution of X_i given $V_i = v$. This is in line with Khan and Tamer (2010), who pointed out the relationship between square integrability of the weight in a moment condition and regular identification.

Two factors determine the variance of the weight: the dispersion the counterfactual distribution relative to the observed one and instruments strength. Naturally, if the counterfactual distribution of regressors has thin tails relative to the observed distribution, the numerator in $\alpha_0 \equiv f^{x*}/f_0^{x|v}$ will be relatively small. Thus, it will be easier to regularly identify the CASF for those cases. However, as discussed in Section 4, it is the conditional distribution of the regressors what turns out to be relevant. Here is where instrument strength comes into play. With strong instruments, $\pi_0(Z_i)$ greatly explains the endogenous variable (T_i). Hence, the conditional distribution of $X_i \equiv (T_i, Z_{1i})$ given $V_i = v$ will be very disperse, as V_i provides relatively little information about the endogenous variable. Thus, $f_0^{x|v}$ will have fat tails, making the denominator of the weight larger.

To better understand the role of instrument strength in identification of counterfactuals, consider a model with a single regressor T_i and a single instrument Z_i. As in Staiger and Stock (1997) and Jun and Pinkse (2012), I model instrument strength by introducing a parameter h and defining

$$T_i = \pi_0(Z_i) + V_i = h\tilde{\pi}_0(Z_i) + V_i.$$

So $h \to 0$ represents that instruments are weak (the authors assume that instruments get weaker when the sample size increases). In this case, we will have that the variance of α_0 is unbounded. Indeed, as instruments become weaker, the conditional distribution of regressors $f_0^{x|v}$ concentrates around v. Thus, for $x \neq v$, the denominator in α_0 will be arbitrarily small. This is stated in the following lemma:

Lemma 5.2. *Consider a reduced-form with a single regressor T_i and a single instrument Z_i: $T_i = h\tilde{\pi}_0(Z_i) + V_i$. Further, assume that both $\tilde{\pi}_0(Z_i)$ and V_i have densities w.r.t. the Lebesgue measure. Moreover, let Z_i and V_i be independent. Then, for α_0 as in equation* (4.5)*,*

$$\mathrm{E}_{P_0}\left[\alpha_0(X_i, V_i)^2\right] \to \infty \text{ as } h \to 0.$$

This result, combined with Theorem 5.1, rules out regular identification in the case of weak instruments. When the variance of Y_i is bounded bellow, the only relevant term in the necessary condition is the ratio α_0, which has infinite variance.

Corollary 5.1. *Under the assumptions of Theorem 5.1 and Lemma 5.2, if (i) $\sigma_Y^2(x, v) \geq C > 0$ and (ii) $h \to 0$, then the CASF is irregularly identified.*

Boundedness of the conditional variance of Y_i, as in the above corollary, disregards its contribution in to necessary condition to identify the CASF. However, Corollary 5.1 still covers meaningful cases. In the binary choice model, the conditional variance of Y_i is

$$\sigma_Y^2(x,v) = F_0(\theta_0(x),v)(1 - F_0(\theta_0(x),v)).$$

Therefore, if (i) the support of X_i is compact, (ii) θ_0 is continuous, and (iii) the support of U_i (conditional on $V_i = v$) is the whole real line; then the conditional variance of Y_i is bounded away from zero. Compact support of the regressors is common in empirical applications and is usually assumed for estimation (see, among others, Imbens and Newey, 2009; Rothe, 2010; Hsu et al., 2020).

Nevertheless, the term σ_Y^2 can help in satisfying the necessary condition for regular identification of the CASF. That is, even if $\mathrm{E}_{P_0}\left[\alpha_0(X_i,V_i)^2\right] = \infty$, the necessary condition will be satisfied if $\sigma_Y^2(x,v)$ tends to zero fast enough. This gives rise to a large array of possibilities, which are studied in Section 6 for normal DGPs. Here, I discuss what determines the shape of σ_Y^2: the support of the index $\theta_0(X_i)$ (relevance of regressors) and the support of the structural error U_i given $V_i = v$ (degree of endogeneity).

Relevance of the regressors: If the support of the index $\theta_0(X_i)$ is large relative to the one of U_i given $V_i = v$, one may expect to find points (x,v) for which $F_0(\theta_0(x),v) = \mathrm{E}_{P_0}\left[Y_i | X_i = x, V_i = v\right]$ is close to zero or one. In this case, it will be easier to regularly identify the CASF. This mainly depends on the range of the function θ_0, that is, in how relevant regressors are in determining the value of Y_i. To illustrate, consider that there is a regressor in X_i, say X_{1i}, with (i) support in the whole real line and (ii) that satisfies $\lim_{|x_1|\to\infty} |\theta_0(x)| = \infty$. That is, X_{1i} has a sizable impact on Y_i (this is the case, for instance, of a linear index with a not null coefficient for X_{1i}). In this case, $\lim_{|x_1|\to\infty} \sigma_Y^2(x,v) = 0$, so the necessary will be easier to satisfy. This finding is similar in spirit to Vytlacil and Yildiz (2007), which require strong exogenous regressors for their identification strategy.

Degree of endogeneity: For the conditional variance of Y_i to be close to zero, it is also advantageous that the regressors are highly endogenous, that is, that U_i and V_i are highly dependent. For instance, consider the case where the conditional expectation of U_i given $V_i = v$ is linear.[4] Then, the law of total variance yields that the expected conditional variance of U_i given $V_i = v$ decreases in the correlation coefficient of (U_i, V_i). Therefore, if the correlation coefficient is high, the distribution of U_i conditional on $V_i = v$ (namely, F_0) will have thinner tails, and thus $F_0(u,v)(1 - F_0(u,v))$ will be closer to zero for smaller u. Intuitively, one can argue that if U_i and V_i are highly dependent, the researcher may learn about U_i by "observing" V_i (since V_i is identified as a first-stage residual).

[4]Recall that Pearson's correlation coefficient only measures the strength of the linear relationship between two variables, and thus, some linear dependence structure must be imposed to analyze its effect.

6 Normal DGPs

In this section, I study how the necessary condition in Theorem 5.1 behaves when the underlying DGP is normal (this excercise is similar to Khan and Tamer, 2010, who study different DGPs to illustrate that the convergence rate varies with them). Under a normal DGP, I am able to jointly analyze α_0^2 and σ_Y^2. I find that the necessary condition for regular identification translates into a condition on the structural parameters of the model. In line with the preceding discussion, these parameters measure instrument strength, the degree of endogeneity and relevance of the regressors.

Being the counterfactual distribution fixed, whether the CASF is regularly identified depends on the underlying DGP. The present exercise illustrates that, in general, the non-parametric model contains both DGPs for which the necessary condition for regular identification is met, and DGPs for which it is not. Hence, when non-parametrically estimating the CASF, the researcher is uncertain about whether it is regularly or irregularly identified. The results from this section provide a way to informally check whether one is in the irregularly identified case.

To simplify the matter, consider that there is a single endogenous variable T_i.[5] The reduced-from equation for T_i includes a single instrument Z_i. I study the DGP emerging from linear θ_0 and π_0. That is, P_0 is the distribution of $(Y_i, T_i, Z_i, U_i, V_i)$ implied by

$$\begin{aligned} Y_i &= \mathbb{1}\left(\theta_{00} + \theta_{10} T_i + U_i \geq 0\right), \\ T_i &= \pi_0 Z_i + V_i, \end{aligned} \tag{6.1}$$

and

$$\begin{pmatrix} Z_i \\ -U_i \\ V_i \end{pmatrix} =^d N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_Z^2 & 0 & 0 \\ 0 & \sigma_U^2 & \sigma_{UV} \\ 0 & \sigma_{UV} & \sigma_V^2 \end{pmatrix} \right), \tag{6.2}$$

where $=^d$ denotes equally distributed. I have centered Z_i, V_i and T_i around zero to ease computations.

In what follows, I denote the variance of T_i with $\sigma_T^2 \equiv \pi_{10}^2 \sigma_Z^2 + \sigma_V^2$. I also introduce the coefficient of determination of the regression of T_i on Z_i:

$$R^2 \equiv \frac{\pi_{10}^2 \sigma_Z^2}{\sigma_T^2} = 1 - \frac{\sigma_V^2}{\sigma_T^2}.$$

The R^2 coefficient of the reduced-form is extremely relevant for regular identification of the CASF. It is, indeed, a measure of instrument strength, which is a determinant of counterfactual identification (see Section 5). The remaining determining factors were the relevance of the regressor in the structural equation and endogeneity of the regressor. To map these

[5] Since there are no exogenous regressors in the structural equation, throughout this section $X_i = T_i$.

into parameters of the model, I introduce the following measures:

$$\Theta \equiv \frac{\sigma_T}{\sigma_U}\theta_{10}, \text{ and } \rho_{UV} \equiv \frac{\sigma_{UV}}{\sigma_U \sigma_V}.$$

The first is a measure of the relevance of T_i in the structural equation. Note that, in the current setup, the ASF is given by

$$\text{ASF}_0(t) = \Phi\left(\frac{\theta_{00} + \theta_{10}t}{\sigma_U}\right).$$

Therefore, Θ is the standard deviation of the argument in the ASF. The parameter does also take into account the sign of θ_{10}, which turns out to be relevant. The second quantity is the coefficient of correlation of (U_i, V_i). It is a measure of the degree of endogeneity of the regressor (T_i), since $\text{Cov}(U_i, T_i) = \text{Cov}(U_i, V_i) \equiv \sigma_{UV}$.

For normal DGPs, one can characterize the tails of α_0^2 and σ_Y^2. Therefore, one can check whether $\alpha_0^2\sigma_Y^2$ is integrable. The third ingredient to study the necessary condition in equation (5.1) is the measure with respect to which the integral is computed: the distribution of (X_i, V_i) under P_0. Before jumping into the main discussion, I briefly introduce an alternative measure. This will help interpreting the results. For any counterfactual distribution f^{t*} for T_i, I can write

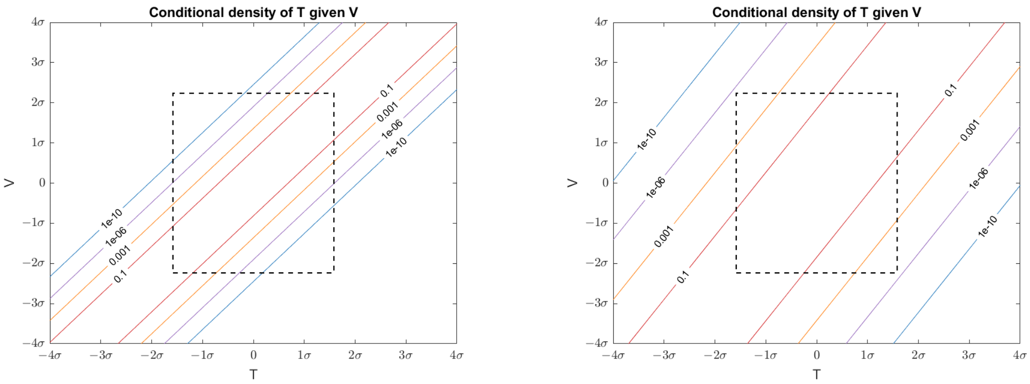
$$\begin{aligned}
\text{E}_{P_0}\left[\alpha_0(T_i, V_i)^2\sigma_Y^2(T_i, V_i)\right] &= \iint \left(\frac{f^{t*}(t)}{f_0^{t|v}(t,v)}\right)^2 \sigma_Y^2(t,v) f_0^{t|v}(t,v) f_0^v(v)dtdv = \\
&= \iint \frac{\sigma_Y^2(t,v)}{f_0^{t|v}(t,v)} f^{t*}(t)^2 f_0^v(v)dtdv = \\
&= \iint \frac{\sigma_Y^2(t,v)}{f_0^{t|v}(t,v)} dm(t,v),
\end{aligned}$$

with $m(A) \equiv \iint_A (f^{t*})^2 f_0^v dtdv$ for any measurable set A. Therefore, regular identification of the CASF depends on the relative tails of σ_Y^2 and $f_0^{t|v}$. The integrating measure depends on the counterfactual distribution. If the counterfactual distribution is very disperse, m will have its mass spread out in the plane. Thus, the tails of $\sigma_Y^2/f_0^{t|v}$ will have to decline faster.

6.1 Variance shift counterfactual

Consider that the goal of the researcher is to estimate the CASF for $T_i =^d N(0, \sigma_{T^*}^2)$. That is, the counterfactual distribution of the endogenous regressor has the same mean but may be more (or less) disperse: a mean preserving spread. For instance, in the empirical application, where T_i represents GDP growth, one may be interested in the effects market integration, which could reduce the variance of T_i. The main discussion is centered around this counterfactual. I later analyze regular identification of the CASF under a mean shift. This turns out to be a particular case of the results for the variance shift.

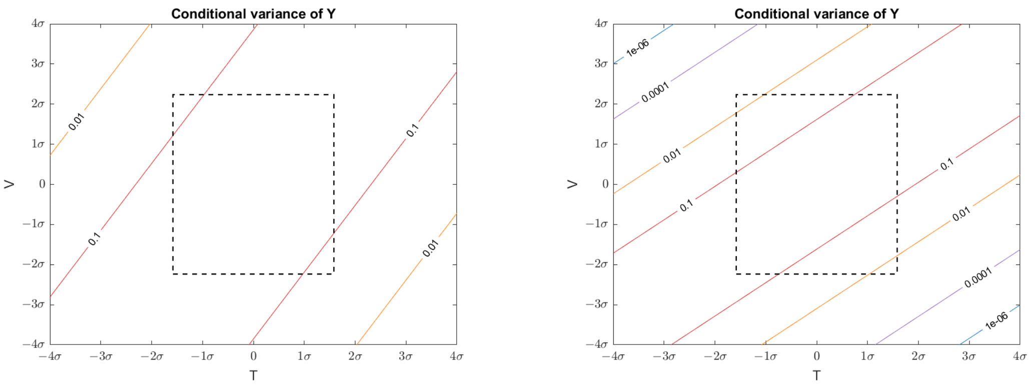
To get a deeper insight about what determines regular identification of the CASF, one can see how the tails of $f_0^{t|v}$ depend on instrument strength. Figure 1 displays the level curves of the conditional density for $R^2 \in \{0.3, 0.6\}$. Clearly, the integral of $\sigma_Y^2/f_0^{t|v}$ with respect to m is larger when instruments are weaker (σ_Y^2 remains unchanged). To quantify how fast $f_0^{t|v}$ declines when compared to the integrating measure m, I plot the region where m encloses 95% of its mass. Recall that the measure m is determined by the counterfactual distribution, that is, by σ_{T*}^2. The region is computed for $\sigma_{T*}^2 = \sigma_T^2$, i.e., the factual distribution of T_i, which may serve as a benchmark. If $R^2 = 0.3$, the conditional density is already bellow 10^{-10} in this region.



(a) $R^2 = 0.3$ **(b)** $R^2 = 0.6$

Figure 1: Level curves of $f_0^{t|v}$ for different values of instrument strength. The dashed region encloses 95% of the mass of the measure m, computed for $\sigma_{T^*}^2 = \sigma_T^2$.

The same exercise can be done for the conditional variance of Y_i. Figure 2 displays the level curves of σ_Y^2 for $\rho_{UV} \in \{0.3, 0.6\}$. The decay of the conditional variance is faster the greater ρ_{UV} is. This may help compensate that $f_0^{t|v}$ converges rapidly to zero, and thus it eases that $\sigma_Y^2/f_0^{t|v}$ is integrable w.r.t. m. That is, if the degree of endogeneity (ρ_{UV}) is larger, it is easier to regularly identify the CASF. The picture of increasing relevance (Θ) is similar to the one in Figure 2: the stronger the regressor, the easier to regularly identify the CASF.



(a) $\rho_{UV} = 0.3$ **(b)** $\rho_{UV} = 0.6$

Figure 2: Level curves of σ_Y^2 for different values of the degree of endogeneity. Relevance is fixed to $\Theta = 0.5$. The dashed region encloses 95% of the mass of the measure m, computed for $\sigma_{T^*}^2 = \sigma_T^2$.

Up to now, I have separately analyzed the two terms that determine regular identification of the CASF: α_0 and σ_Y^2. For normal DGPs, I can establish a condition under which $\alpha_0^2\sigma_Y^2$ is integrable:

Proposition 6.1. *Consider a normal P_0, as described by equations* (6.1) *and* (6.2). *Then, for the counterfactual $T_i =^d N(0, \sigma_{T^*}^2)$, $\mathrm{E}_{P_0}\left[\alpha_0(T_i, V_i)^2\sigma_Y^2(T_i, V_i)\right] < \infty$ is equivalent to*

$$\left[1 - \Theta^2 + 2\sqrt{1-R^2}\Theta\left(\sqrt{1-R^2}\Theta - \rho_{UV}\right)\right]\frac{\sigma_{T^*}^2}{\sigma_T^2} < 2\left[1 - (2-\rho_{UV}^2)(1-R^2)\right]. \qquad (6.3)$$

According to the above result, regular identification of the CASF in the counterfactual scenario given by $N(0, \sigma_{T^*}^2)$ is exclusively determined by four parameters: R^2 (instrument strength), Θ (relevance), ρ_{UV} (endogeneity), and the ratio of variances ($\sigma_{T^*}^2/\sigma_T^2$). This last parameter measures the relative dispersion of the counterfactual when compared with the observed distribution.

Importantly, the above condition can be checked after fitting the model following Rivers and Vuong (1988). One can get estimates of $(\sigma_T^2, R^2, \rho_{UV}, \Theta)$ from the two-step procedure described on the paper: (i) fit a linear model for T_i on Z_i and obtain the OLS residuals $\hat{V}_i$, then (ii) fit a probit model for Y_i on T_i and $\hat{V}_i$. Hence, condition (6.3) can be used as an informal test for regular identification of the CASF. If the condition is not met, the researcher may suspect that the CASF is irregularly identified for the counterfactual at hand. Providing a formal test for regular identification is out of the scope of this paper.

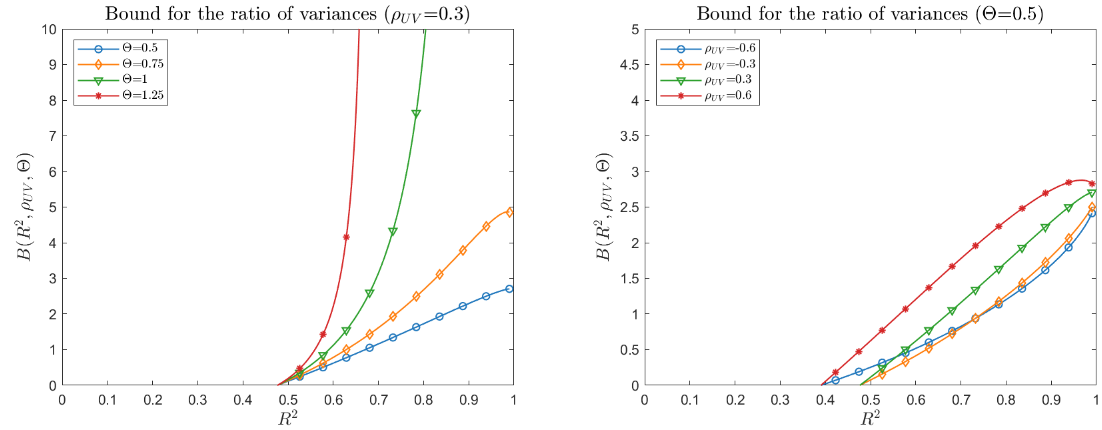
Condition (6.3) provides an upper bound on how big the variance of the counterfactual distribution can be, relative to the observed variance, so that the CASF may be regularly identified (see Corollary 6.1 bellow). If the variance of the counterfactual distribution is large, the counterfactual T_i is much more scattered than the observed random variable. This means that draws that would usually occur in the counterfactual world are rare to see in the actual data. In this situation, the data at hand will be uninformative when it comes to estimate the ASF for these extreme points. Since the ASF is needed to compute the CASF, the parameter is therefore irregularly identified.

Corollary 6.1. *Under the conditions of Proposition 6.1, the CASF is irregularly identified if $\sigma_{T^*}^2/\sigma_T^2 \geq B(R^2, \rho_{UV}, \Theta)$, where*

$$B(R^2, \rho_{UV}, \Theta) \equiv \begin{cases} \frac{1-(2-\rho_{UV}^2)(1-R^2)}{H(R^2,\rho_{UV},\Theta)} & \text{if } H(R^2, \rho_{UV}, \Theta) > 0 \\ \infty & \text{if } H(R^2, \rho_{UV}, \Theta) \leq 0 \end{cases},$$

and $H(R^2, \rho_{UV}, \Theta) \equiv 1 - \Theta^2 + 2\sqrt{1-R^2}\Theta\left(\sqrt{1-R^2}\Theta - \rho_{UV}\right)$ is the LHS of condition (6.3).

Interestingly, there are certain combinations of the parameters (R^2, ρ_{UV}, Θ) for which one may regularly identify the CASF for any counterfactual variance σ_{T*}^2. This scenario is the most favorable one for counterfactual analysis. It happens when the parameters take large values: the LHS of condition (6.3) is negative when R^2 and $|\rho_{UV}|$ are close to 1, and $|\Theta|$ is large (see the proof of Lemma A.2 in the Appendix). This confirms the results about the non-parametric determinants of regular identification of the CASF (see Section 5). Regular identification of the CASF is easier with (i) strong instruments (R^2), (ii) relevant regressors (Θ), and (iii) highly endogenous regressors (ρ_{UV}). Figure 3 backs up these findings by showing how large the counterfactual variance σ^2_{T*} can be, relative to the observed one, for different values of (R^2, ρ_{UV}, Θ).



(a) **(b)**

Figure 3: Bound on σ^2_{T*}/σ^2_T for the CASF to be regularly identified. In the left figure, the bound is infinity (no bound) for $R^2 \in [0.91, 1)$ when $(\rho_{UV}, \Theta) = (0.3, 1)$, and $R^2 \in [0.685, 1)$ when $(\rho_{UV}, \Theta) = (0.3, 1.25)$.

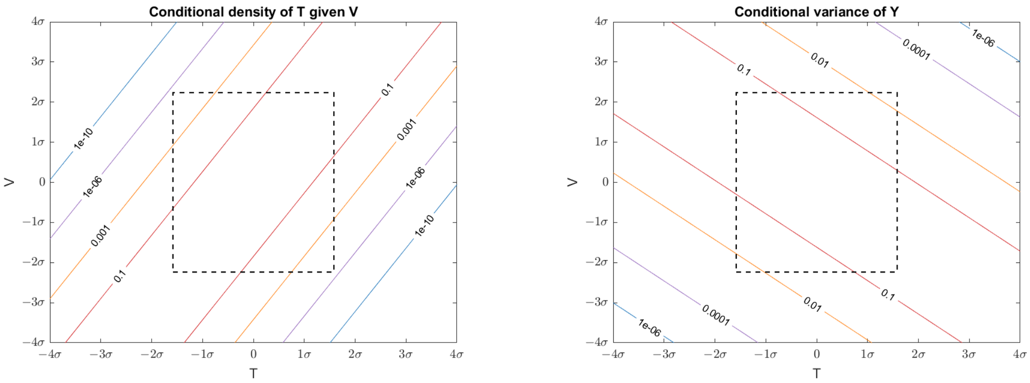
Figure 3 does also disclose a relevant fact about the impact of instrument strength: there is a value of R^2 bellow which regular identification of the CASF is not possible. This is an immediate follow-up of Proposition 6.1, which I highlight here:

Corollary 6.2. *Under the conditions of Proposition 6.1, the CASF is irregularly identified if*

$$R^2 \leq \underline{R}^2 \equiv 1 - \frac{1}{2 - \rho_{UV}^2}.$$

This result imposes a lower bound on instrument strength: if the instrument at hand is not strong enough, one will not be able to regularly identify the CASF. The threshold ranges from 1/2, when $\rho_{UV} = 0$, to 0, when $|\rho_{UV}| = 1$. Researchers can evaluate the above condition as an informal check for regular identification of the CASF (see the empirical application in Section 7 for an illustration).

In the left display of Figure 3, I show the consequences of Θ and ρ_{UV} having opposite sign. When the parameters have different signs, the bound on the variance ratio is smaller. An explanation is displayed in Figure 4. The tails of the conditional distribution of T_i given $V_i = v$ always decrease in the direction of the $(1, -1)$ vector (NW to SE). When Θ and ρ_{UV} have different signs, the tails of the conditional variance σ^2_Y decrease in the $(1, 1)$ direction (SW to NE). Hence, σ^2_Y is not able to compensate for $f_0^{t|v}$ being small: when the later is close to zero (e.g., at point $(2\sigma, -2\sigma)$), the former is far away from zero. This does not happen when Θ and ρ_{UV} have the same sign (see Figures 1 and 2).



(a) **(b)**

Figure 4: Level curves of $f_0^{t|v}$ (left) and σ_Y^2 (right) when $\Theta = 0.5$ and $\rho_{UV} = -0.6$ have different signs. Instrument strength is fixed to $R^2 = 0.6$. The dashed region encloses 95% of the mass of the measure m, computed for $\sigma_{T^*}^2 = \sigma_T^2$.

6.2 Mean shift counterfactual

In this section I study the case where the researcher is interested in estimating CASF for $T_i =^d N(\mu_{T^*}, \sigma_T^2)$. The location shift case is of especial interest, as many policies could be framed into the setup (such as lump-sum transfers or cost reductions). In the empirical application, one can think that international aid shifts upwards the growth rate of Sub-Saharan countries.

The mean shift counterfactual is a special case of the variance shift studied above. Indeed, the necessary condition for regular identification of the CASF is a condition imposed on the tails of the counterfactual distribution. These are not affected by the mean shift. Nevertheless, to regularly identify the CASF in a mean shift counterfactual, one should still be able to identify a counterfactual variance as large as the observed one. The resulting condition is thus a particular case of condition (6.3), where $\sigma_{T^*}^2/\sigma_T^2$ is set to 1.

Proposition 6.2. *Consider a normal P_0, as described by equations* (6.1) *and* (6.2)*. Then, for the counterfactual $T_i =^d N(\mu_{T^*}, \sigma_T^2)$, $\mathrm{E}_{P_0}\left[\alpha_0(T_i, V_i)^2 \sigma_Y^2(T_i, V_i)\right] < \infty$ is equivalent to*

$$1 - \Theta^2 + 2\sqrt{1-R^2}\,\Theta\left(\sqrt{1-R^2}\,\Theta - \rho_{UV}\right) < 2\left[1 - (2-\rho_{UV}^2)(1-R^2)\right].$$

The above result highlights that the model may not be able to identify the CASF for a counterfactual distribution equal to the observed distribution of the regressors. This is a consequence of having endogenous regressors. In the control function approach, identification of the ASF is achieved by exogenous moves of the regressors (those conditional on $V_i = v$). The conditional variance of T_i given $V_i = v$ depends on instrument strength. Therefore, if instruments are weak (R^2 close to 0), one will not be able to regularly identify a counterfactual distribution with the same dispersion as the factual one. In this case, the CASF will be irregularly identified for every counterfactual mean μ_{T^*}. This is a strong result, as the effect of any policy that alters the mean of the regressors, no matter how small this change is, will not be estimable at the parametric rate.

7 An empirical application

This section illustrates the ideas discussed above with empirical data. In particular, I study the impact of economic growth on the likelihood of civil conflict (a detailed overview of the literature can be found in Blattman and Miguel, 2010). Economic growth (per capita GDP growth) is likely endogenous in an equation for the presence of civil conflict (more than 25 battle-related deaths) in a country at a given year. Thus, I follow Miguel et al. (2004) and use variation in rainfall as an instrument.

The econometric model at hand is:

$$\begin{aligned}\text{Conflict}_i &= \mathbb{1}\left(\theta_0(\Delta\text{GDP}_i) + U_i \geq 0\right),\\ \Delta\text{GDP}_i &= \pi_0(\Delta\text{Rainfall}_i) + V_i,\end{aligned}$$

where the observation unit i represents a country-year pair. The sample includes all the Sub-Saharan countries with available data and spans from 1981 to 1999. Data on conflict comes from UCDP/PRIO Armed Conflict Dataset, rainfall data comes from the GPCP Combined Precipitation Data Set and the World Bank is the source for per capita GDP growth.

To keep the model close to the discussion in Section 6, I simplify Miguel et al. (2004)'s specification: I do not include lagged GDP growth nor other potential determinants of conflict (measures for democracy, orography, religious or ethnolinguistic fractionalization) in the structural equation. Ultimately, the goal of the application is to study how easy it is to perform counterfactual analysis with this data. Nevertheless, there is conflicting evidence about the impact of democracy and fractionalization measures on civil conflict prevalence (cf. Elbadawi and Sambanis, 2002; Miguel et al., 2004).

As discussed throughout the paper, the CASF can be used to answers a variety of policy relevant questions in this setup. It remains to see, however, whether the CASF is regularly or irregularly identified. Miguel et al. (2004) argue that rainfall is a valid instrument for GDP growth in economies where the agricultural sector is large and irrigation is not widely extended, which is the case of many Sub-Saharan countries. But it may be a weak instrument. Indeed, a linear first-stage (i.e., $\pi_0(z) = \pi_{00} + \pi_{10}z$) yields an R^2 as low as 0.0244 (Miguel et al., 2004, report $R^2 = 0.02$ in a linear specification which further includes lagged rainfall). With this value, the CASF will probably be irregularly identified, if it is identified at all.

A non-parametric first-stage also points out to irregular identification of the CASF. Figure 5 depicts the conditional density of per capita GDP growth given the control variable and the conditional variance of the presence of conflict given per capita GDP growth and the control variable. In panel (a) one can see that the control variable greatly explains per capita GDP growth, as rainfall growth is a weak instrument. Therefore, the tails of the conditional density decay rapidly: they are already bellow 10^{-10} in the region where m puts

95% of its mass. Panel (b) confirms that the conditional variance will not compensate for this fast decay. Hence, the necessary condition for regular identification of the CASF is hardly satisfied for this data.

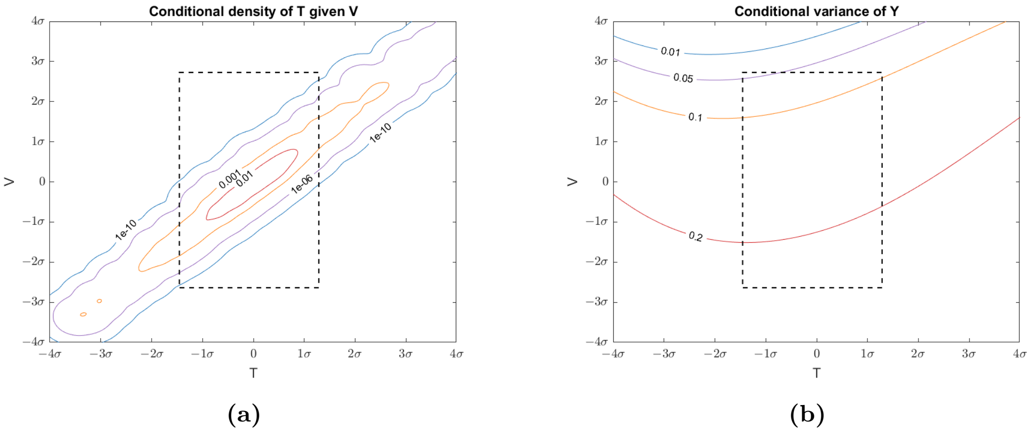


Figure 5: Level curves of (a) the conditional density of per capita GDP growth (T_i) given the control variable (V_i) and (b) the conditional variance of the presence of conflict (Y_i) given per capita GDP growth and the control variable. The control variable is the residual from a first-stage non-parametric regression of per capita GDP growth on rainfall growth (sample size of 688). The dashed region encloses 95% of the mass of the measure m, computed for $f^{t*} = \hat{f}^t$, a Gaussian kernel estimator of the density of per capita GDP growth.

One can use the results of Section 6 as an informal check for regular identification of the CASF in this model. The first-stage R^2 coefficient is 0.0244, which will probably not exceed the threshold in Corollary 6.2. To confirm it, the researcher can fit a normal model using Rivers and Vuong (1988)'s methodology. In this case, the values of the remaining parameters are $\rho_{UV} = -0.0427$ and $\Theta = -0.0745$. Hence, the R^2 coefficient is well bellow the threshold of $\underline{R}^2 = 0.4995$ and the CASF is probably irregularly identified.

8 Conclusion

This paper studies regular identification of the CASF. In the binary choice model, the parameter measures the rate of "successes" when the regressor's distribution is set to a counterfactual. Thus, the CASF is useful in measuring policy effects and performing decomposition analysis. I find a necessary condition for regular identification of the CASF. There are several counterfactual scenarios for which the condition is not met, so the CASF is irregularly identified. Furthermore, being the counterfactual distribution fixed, I show that regular identification of the CASF is easier with strong instruments, relevant regressors and highly endogenous regressors.

I characterize this dependence for normal DGPs. There, the necessary condition for regular identification of the CASF translates into a condition for the structural parameters. This helps understanding how instrument strength, relevance and the degree of endogeneity of the regressors affect counterfactual identification. Moreover, this analysis provides an informal check for regular identification of the CASF. For instance, I show that there exist a threshold for instrument strength bellow which regular identification of the CASF is not possible.

The key feature of (regular) identification of the CASF is that the parameter may be written as a weighted average, the weight given by the likelihood ratio between the counterfactual density and the conditional density of regressors given control variables. It is thus expected that the points discussed in this work can be extended to other parameters and settings with a similar feature. In particular, any moment of the outcome in a counterfactual scenario (e.g., variance, quantiles) could be subject to the same discussion. Moreover, the ideas of this paper could be employed in other models with endogenous regressors where identification is achieved by a control function assumption (e.g., Imbens and Newey, 2009).

9 Acknowledgments

I am deeply thankful to Juan Carlos Escanciano for his advise and guidance. I thank Miguel A. Delgado, Juan J. Dolado and Joël Terschuur, the editor, the associate editor and three anonymous referees for their helpful comments. I do also acknowledge the importance of the feedback from all the participants at seminars at the Universidad Carlos III de Madrid and Universität Mannheim.

References

Abramowitz, M. and Stegun, L. (1972). *Handbook of Mathematical Functions*. Dover Publications Inc.

Ao, W., Calonico, S., and Lee, Y.-Y. (2021). Multivalued treatments and decomposition analysis: An application to the wia program. *Journal of Business & Economic Statistics*, 39(1):358–371.

Bickel, P., Klaassen, C., Ritov, Y., and Wellner, J. (1998). *Efficient and adaptive estimation for semiparametric models*. Springer-Verlag.

Blattman, C. and Miguel, E. (2010). Civil war. *Journal of Economic literature*, 48(1):3–57.

Blinder, A. S. (1973). Wage Discrimination: Reduced Form and Structural Estimates. *The Journal of Human Resources*, 8(4):436–455.

Blundell, R. W. and Powell, J. L. (2003). Endogeneity in nonparametric and semiparametric regression models. In *Econometric society monographs*, 36.

Blundell, R. W. and Powell, J. L. (2004). Endogeneity in Semiparametric Binary Response Models. *The Review of Economic Studies*, 71(3):655–679.

Carneiro, P., Heckman, J. J., and Vytlacil, E. (2010). Evaluating marginal policy changes and the average effect of treatment for individuals at the margin. *Econometrica*, 78(1):377–394.

Chamberlain, G. (1986). Asymptotic efficiency in semi-parametric models with censoring. *Journal of Econometrics*, 32(2):189–218.

Chernozhukov, V., Fernández-Val, I., and Melly, B. (2013). Inference on counterfactual distributions. *Econometrica*, 81(6):2205–2268.

DiNardo, J., Fortin, N. M., and Lemieux, T. (1996). Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach. *Econometrica*, 64(5):1001–1044.

Elbadawi, I. and Sambanis, N. (2002). How much war will we see? explaining the prevalence of civil war. *Journal of conflict resolution*, 46(3):307–334.

Escanciano, J. C. (2021). Semiparametric identification and Fisher information. *Econometric Theory*, page 1–38.

Fortin, N., Lemieux, T., and Firpo, S. (2011). Decomposition Methods in Economics. In *Handbook of Labor Economics*, volume 4. Elsevier.

Hahn, J. and Ridder, G. (2013). Asymptotic variance of semiparametric estimators with generated regressors. *Econometrica*, 81(1):315–340.

Hahn, J. and Ridder, G. (2019). Three-stage semi-parametric inference: Control variables and differentiability. *Journal of econometrics*, 211(1):262–293.

Heckman, J. J. and Vytlacil, E. (2005). Structural equations, treatment effects, and econometric policy evaluation. *Econometrica*, 73(3):669–738.

Hirano, K. and Imbens, G. W. (2004). The propensity score with continuous treatments. In *Applied Bayesian modeling and causal inference from incomplete-data perspectives.*

Hsu, Y.-C., Lai, T.-C., and Lieli, R. P. (2020). Counterfactual treatment effects: Estimation and inference. *Journal of Business & Economic Statistics*, pages 1–16.

Imbens, G. W. and Newey, W. K. (2009). Identification and estimation of triangular simultaneous equations models without additivity. *Econometrica*, 77(5):1481–1512.

Jun, S. J. and Pinkse, J. (2012). Testing under weak identification with conditional moment restrictions. *Econometric Theory*, 28(6):1229–1282.

Khan, S. and Tamer, E. (2010). Irregular Identification, Support Conditions, and Inverse Weight Estimation. *Econometrica*, 78(6):2021–2042.

Komunjer, I. and Vuong, Q. (2010). Semiparametric efficiency bound in time-series models for conditional quantiles. *Econometric Theory*, 26(2):383–405.

Lewbel, A. (1997). Semiparametric estimation of location and other discrete choice moments. *Econometric Theory*, 13(1):32–51.

Lewbel, A., Dong, Y., and Yang, T. T. (2012). Comparing features of convenient estimators for binary choice models with endogenous regressors. *Canadian Journal of Economics/Revue canadienne d'économique*, 45(3):809–829.

Lin, W. and Wooldridge, J. M. (2015). On different approaches to obtaining partial effects in binary response models with endogenous regressors. *Economics Letters*, 134:58–61.

Matzkin, R. L. (1992). Nonparametric and distribution-free estimation of the binary threshold crossing and the binary choice models. *Econometrica*, 60(2):239–270.

Miguel, E., Satyanath, S., and Sergenti, E. (2004). Economic shocks and civil conflict: An instrumental variables approach. *Journal of political Economy*, 112(4):725–753.

Newey, W. K. (1990). Semiparametric Efficiency Bounds. *Journal of Applied Econometrics*, 5(2):99–135.

Newey, W. K. (1994a). The asymptotic variance of semiparametric estimators. *Econometrica: Journal of the Econometric Society*, pages 1349–1382.

Newey, W. K. (1994b). Kernel Estimation of Partial Means and a General Variance Estimator. *Econometric Theory*, 10(2):1–21.

Oaxaca, R. (1973). Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review*, 14(3):693–709.

Rivers, D. and Vuong, Q. H. (1988). Limited information estimators and exogeneity tests for simultaneous probit models. *Journal of econometrics*, 39(3):347–366.

Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55.

Rothe, C. (2009). Semiparametric estimation of binary response models with endogenous regressors. *Journal of Econometrics*, 153(1):51–64.

Rothe, C. (2010). Nonparametric estimation of distributional policy effects. *Journal of Econometrics*, 155(1):56–70.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5):688.

Shao, J. (2003). *Mathematical Statistics*. Springer Texts in Statistics. Springer-Verlag, New York, 2 edition.

Staiger, D. and Stock, J. (1997). Instrumental variables regression with weak instruments. *Econometrica*, 65(3):557–586.

Stock, J. H. (1989). Nonparametric policy analysis. *Journal of the American Statistical Association*, 84(406):567–575.

Stock, J. H. (1991). Nonparametric policy analysis: an application to estimating hazardous waste cleanup benefits. In *Nonparametric and Semiparametric Methods in Econometrics and Statistics*. Cambridge University Press.

van der Vaart, A. W. (2000). *Asymptotic Statistics*. Cambridge University Press, Cambridge.

Vytlacil, E. and Yildiz, N. (2007). Dummy endogenous variables in weakly separable models. *Econometrica*, 75(3):757–779.

Zimmer, R. J. (1990). *Essential results of functional analysis*. University of Chicago Press.

Appendix A Proofs of the results

Proof of Proposition 4.1.

The proof follows closely the discussion in the text (Section 4). The main idea is that the conditional expectation $K_0(x,v) \equiv \mathrm{E}_{P_0}[Y_i|X_i = x, V_i = v]$ is only identified in the support of (X_i, V_i). That is, the data is uninformative about $K_0(x,v)$ if $(x,v) \in \sup(X_i,V_i)^c$, the complementary of the support of (X_i,V_i). To find the ASF at a point $x \notin \mathcal{S}$, one must integrate along some (x,v) outside the support of (X_i,V_i). Thus, the ASF will not be identified at that point. If this point x is, in turn, in the support of the counterfactual distribution for X_i, then the CASF will not be identified. Bellow, I formalize this procedure. First, to simplify notation, I call $\mathbb{X}^* \equiv \sup_{P^{x*}}(X_i)$ and $\mathbb{V} \equiv \sup(V_i)$.

One can modify the function K_0 in $A \subset \sup(X_i,V_i)^c$ in the following way:

$$K_0'(x,v) = k', \forall (x,v) \in A, \text{ and } K_0'(x,v) = K_0(x,v), \forall (x,v) \in A^c,$$

and have that K_0 and K_0' are observationally equivalent.

I now show that the ASF is identified only for the points in $\mathcal{S}$. If $x \notin \mathcal{S}$, there exist a set $B \subseteq \mathbb{V}$ with $P_0^v(B) > 0$ and for those $v \in B$, $x \notin \sup(X_i|V_i = v)$. This means that $A \equiv \{x\} \times B \subset \sup(X_i,V_i)^c$. That is, following equation (4.1),

$$\begin{aligned}\mathrm{ASF}_0'(x) = \int K_0'(x,v)dP_0^v(v) &= \int_{\mathbb{V}\cap B^c} K_0(x,v)dP_0^v(v) + \int_{\mathbb{V}\cap B} k' dP_0^v(v) = \\ &= \mathrm{ASF}_0(x) + \int_B \left(k' - K_0(x,v)\right) dP_0^v(v)\end{aligned}$$

and $\mathrm{ASF}_0(x)$ are observationally equivalent. Since $P_0^v(B) > 0$, one can find k' such that $\mathrm{ASF}_0(x) \neq \mathrm{ASF}_0'(x)$, so $\mathrm{ASF}_0(x)$ is not identified. Clearly, one cannot conduct the above argument if $x \in \mathcal{S}$.

To conclude, if $\mathbb{X}^* \nsubseteq \mathcal{S}$, since

$$\int \mathrm{ASF}(x)dP^{x*}(x) = \int_{\mathbb{X}^*\cap\mathcal{S}} \mathrm{ASF}(x)dP^{x*}(x) + \int_{\mathbb{X}^*\cap\mathcal{S}^c} \mathrm{ASF}(x)dP^{x*}(x),$$

one could choose observationally equivalent ASFs, i.e., $\mathrm{ASF}_0(x)$ and $\mathrm{ASF}_0'(x)$ for all the $x \in \mathbb{X}^* \cap \mathcal{S}^c$, leading to observationally equivalent CASFs. This is not possible if $\mathbb{X}^* \subseteq \mathcal{S}$ $\square$

Proof of Proposition 4.2.

Assumption 4.2 implies that V_i has marginal pdf $f_0^v = \int f_0^{xv} d\nu^x$ (Shao, 2003, p. 22) and that $X_i|V_i = v$ has conditional pdf $f_0^{x|v}(x,v) = f_0^{xv}(x,v)/f_0^v(v)$ (Shao, 2003, p. 39). I now proof that, under Assumption 4.1, $P_0^v \times P^{x*}$ is absolutely continuous with respect to P_0^{xv}. Let A be a measurable set with $P_0^{xv}(A) = 0$. Then,

$$\iint_A f_0^{xv}(x,v)d\nu^v d\nu^x = 0,$$

which implies that if either $f_0^{xv} = 0$ in A or $\nu^v \times \nu^x(A) = 0$. Since $f_0^{xv} = f_0^{x|v} f_0^v$, it must be that $f_0^{x|v} = 0$ or $f_0^v = 0$ if $f_0^{xv} = 0$. Finally, note that Assumption 4.1 implies that $f_0^{x|v} = 0 \Rightarrow f^{x*} = 0$. In any case, ones has that

$$(P_0^v \times P^{x*})(A) = \iint_A f_0^v(v) f^{x*}(x) d\nu^v d\nu^x = 0.$$

Therefore, a change of measure is allowed in equation (4.3):

$$\begin{aligned}
\gamma_0^* &= \iint \mathrm{E}_{P_0}\left[Y_i | X_i = x, V_i = v\right] \frac{f_0^v(v) f^{x*}(x)}{f_0^{xv}(x, v)} f_0^{xv}(x, v) d\nu^v d\nu^x = \\
&= \mathrm{E}_{P_0}\left[\mathrm{E}_{P_0}\left[Y_i | X_i, V_i\right] \frac{f_0^v(V_i) f^{x*}(X_i)}{f_0^{xv}(X_i, V_i)}\right] = \mathrm{E}_{P_0}\left[Y_i \frac{f_0^v(V_i) f^{x*}(X_i)}{f_0^{xv}(X_i, V_i)}\right] = \\
&= \mathrm{E}_{P_0}\left[Y_i \alpha_0(X_i, V_i)\right].
\end{aligned}$$

□

Proof of Lemma 5.1.
A one-dimensional parameter γ_0^*, as is the CASF, is given by a functional of the distributions in the model. That is, $\gamma_0^* = \psi(P_0)$ for $\psi : \mathcal{P} \to \mathbb{R}$, which we consider to be differentiable (see van der Vaart, 2000; Newey, 1990). A parametric submodel, P_ϕ, is indexed by a function ϕ mapping an interval $[0, T_\phi]$ to the space of distributions of $(Y_i, T_i, Z_{1i}, Z_{2i}, U_i, V_i)$. I collect all the submodels under consideration in the set

$$\Phi \equiv \{\phi : \phi\left([0, T_\phi]\right) \subset \mathcal{P}, \phi(0) = P_0, \text{ and } P_\phi \text{ is differentiable in quadratic mean}\}.$$

Φ' is defined as above but replacing $\mathcal{P}$ by $\mathcal{P}'$. Let $\tilde{\phi} \in \Phi'$. Since, $\tilde{\phi}\left([0, T_\phi]\right) \subset \mathcal{P}' \subset \mathcal{P}$, we have that $\tilde{\phi} \in \Phi$. Therefore

$$\sup_{\phi \in \Phi} \dot{\psi}_{0,\phi}' \, \mathrm{E}_{P_0}\left[S_{0,\phi}^2\right]^{-1} \dot{\psi}_{0,\phi} \geq \dot{\psi}_{0,\tilde{\phi}}' \, \mathrm{E}_{P_0}\left[S_{0,\tilde{\phi}}^2\right]^{-1} \dot{\psi}_{0,\tilde{\phi}},$$

where $\dot{\psi}_{0,\phi}$ denotes the derivative of ψ along the parametric submodel P_ϕ evaluated at P_0. $S_{0,\phi}$ is the score of the model along the parametric submodel P_ϕ. Since this hold for any $\phi \in \Phi'$, I can conclude that

$$\sup_{\phi \in \Phi} \dot{\psi}_{0,\phi}' \, \mathrm{E}_{P_0}\left[S_{0,\phi}^2\right]^{-1} \dot{\psi}_{0,\phi} \geq \sup_{\phi \in \Phi'} \dot{\psi}_{0,\phi}' \, \mathrm{E}_{P_0}\left[S_{0,\phi}^2\right]^{-1} \dot{\psi}_{0,\phi}.$$

This concludes the proof, as the LHS is the efficiency bound under model $\mathcal{P}$ and the RHS is the bound under $\mathcal{P}'$ (see van der Vaart, 2000, Ch. 25). □

Proof of Theorem 5.1.
I begin the proof by imposing additional restrictions to the model, i.e., defining the submodel $\mathcal{P}'$ at which I will find the information bound for the CASF. By Lemma 5.1, the information bound for model $\mathcal{P}$ in Assumption 3.1 is no smaller than the one for the submodel $\mathcal{P}'$. The

more restricted model satisfies Assumptions 3.1, 4.1, 4.2 and

Assumption A.1.

a. π_0 is a known function, i.e., $\pi \in \{\pi_0\}$.

b. The probability distribution P is such that the distribution of (X_i, V_i) is known, i.e., $f^{xv} \in \{f_0^{xv}\}$.

c. The conditional distribution of U_i given $V_i = v$ has density $f^{u|v}(u|v) = \frac{\partial F}{\partial u}(u, v)$, which is bounded.

Assumption A.1.a states that $V_i = T_i - \pi_0(Z_i)$ is now an observed random variable. I can thus directly work with the density of (Y_i, X_i, V_i). To find the tangent space of the model, I parametrize this density by an infinite dimensional parameter (see van der Vaart, 2000, Ch. 25). Under the control function assumption:

$$\begin{aligned}P(Y_i = 1|X_i = x, V_i = v) = P(Y_i(x)|X_i = x, V_i = v) = P(Y_i(x)|V_i = v) = \\ = P(-U_i \leq \theta_0(x)|V_i = v) = F(\theta_0(x), v),\end{aligned}$$

where $F(u, v) \equiv P(-U_i \leq u|V_i = v)$ is the conditional cdf of (minus) the strutural error given the control variable. Thus, under the above assumptions the density of (Y_i, X_i, V_i) is given by

$$p(y, x, v) = F(\theta(x), v)^y \left(1 - F(\theta(x), v)\right)^{1-y} f_0^{xv}(x, v)$$

w.r.t. the measure $\mu \equiv c \times \nu^x \times \nu^v$. That is, the unknown parameters of the model are (θ, F).

Consider now a one-dimensional differentiable submodel $\tau \mapsto (\theta_\tau, F_\tau)$ with density $p_\tau(y, x, v) = F_\tau(\theta_\tau(x), v)^y \left(1 - F_\tau(\theta_\tau(x), v)\right)^{1-y} f_0^{xv}(x, v)$. The score of an arbitrary submodel is

$$\begin{aligned}S_0 &= y\frac{\dot{F}_0(\theta_0(x), v) + f_0^{u|v}(\theta_0(x)|v)\dot{\theta}_0(x)}{F_0(\theta_0(x), v)} - (1-y)\frac{\dot{F}_0(\theta_0(x), v) + f_0^{u|v}(\theta_0(x)|v)\dot{\theta}_0(x)}{1 - F_0(\theta_0(x), v)} = \\ &= \left(\dot{F}_0(\theta_0(x), v) + f_0^{u|v}(\theta_0(x)|v)\dot{\theta}_0(x)\right)\frac{y - F_0(\theta_0(x), v)}{F_0(\theta_0(x), v)(1 - F_0(\theta_0(x), v))},\end{aligned}$$

where

$$\dot{F}_0(u, v) \equiv \left.\frac{\partial F_\tau(u, v)}{\partial \tau}\right|_{\tau=0} \text{ and } \dot{\theta}_0(x) \equiv \left.\frac{\partial \theta_\tau(x)}{\partial \tau}\right|_{\tau=0}$$

are the derivatives of F_τ and θ_τ along the submodel. I now proceed to show that there is enough flexibility in the submodels considered so that the term $\dot{F}_0 + f_0^{u|v}\dot{\theta}_0$ spans all square-integrable functions of (x, v).

First, Chamberlain (1986) showed that $\dot{F}_0$ are dense in $L_2(P_0)$ (see also Newey, 1990). That is, denoting the index by $\iota \equiv \theta_0(x)$, this term spans any square-integrable function $E(\iota, v)$. Consider now two polynomials $q_1(v)$ and $q_2(x)$ that are zero outside compact sets. Let $Q \equiv \sup|q_1(v)|$. By Assumption A.1.c, $E(\iota, v) \equiv (1 + Q + q_1(v))^{-1} f_0^{u|v}(\iota|v)$ is square-integrable. Moreover, considering the submodel $\theta_\tau(x) \equiv \theta_0(x) + \tau q_2(x)$, the term $\dot{F}_0 + f_0^{u|v}\dot{\theta}_0$

is arbitrarily close to

$$\frac{f_0^{u|v}(\iota|v)}{1+Q+q_1(v)} + f_0^{u|v}(\iota|v)q_2(x) = \frac{f_0^{u|v}(\iota|v)}{1+Q+q_1(v)}\left((1+Q)q_2(x) + q_1(v)q_2(x)\right).$$

The term $(1+Q)q_2+q_1q_2$ contains all polynomials in (x, v) that are zero outside a compact set. The Stone-Weierstrass theorem (see Theorem A.8 and Example A.9.a in Zimmer, 1990) states that these are dense in the space of continuous functions with the same compact support. To conclude, recall that the space of continuous functions with compact support is dense in $L_2(P_0)$ (see Proposition A.10 in Zimmer, 1990). That is, the term $(1+Q)q_2+q_1q_2$ is arbitrarily close to any function $D(x, v) \in L_2(P_0)$. Absorbing all the remaining terms into D and calling $\varepsilon \equiv y - F_0(\theta_0(x), v)$, the tangent space is thus

$$\left\{D(x, v)\varepsilon : \mathrm{E}_{P_0}\left[D(X_i, V_i)^2\right] < \infty\right\}.$$

Assumption A.1.b does also imply that the weight α_0 in the definition of the parameter of interest,

$$\gamma_0^* = \mathrm{E}_{P_0}\left[Y_i\alpha_0(X_i, V_i)\right] = \int y\alpha_0(x, v)p_0(y, x, v)d\mu,$$

is a known function of (x, v). Therefore, the weight remains unchanged along a parametric submodel. The parameter along the submodel is given by $\gamma_\tau^* = \int y\alpha_0(x, v)p_\tau(y, x, v)d\mu$. To show that γ_0^* is a differentiable parameter (Newey, 1990), differentiation under the integral sign along a differentiable submodel must be allowed. Thus, I introduce a regularity assumption in the restricted model given by $\mathcal{P}'$. Define $S_\tau \equiv p_\tau^{-1}\frac{\partial p_\tau}{\partial \tau}$, so that the score along the submodel is given by $S_0 = S_\tau|_{\tau=0}$. I assume that

Assumption A.2. The expectation of $Y_i\alpha_0(X_i, V_i)S_\tau$ along a differentiable parametric submodel, which is given by $\int y\alpha_0(x, v)S_\tau p_\tau d\mu$, exists and is continuous at $\tau = 0$.

Under the above assumption, Lemma B.1 from Zimmer (1990) ensures that differentiation under the integral sign is allowed. Therefore,

$$\begin{aligned}\left.\frac{\partial\gamma_\tau^*}{\partial\tau}\right|_{\tau=0} &= \int y\alpha_0(x, v)\left.\frac{\partial p_\tau}{\partial\tau}\right|_{\tau=0} d\mu = \int y\alpha_0(x, v)S_0p_0d\mu = \\ &= \mathrm{E}_{P_0}\left[Y_i\alpha_0(X_i, V_i)S_0\right].\end{aligned}$$

Hence, one of the Riesz representers of the derivative is $\psi_0 = y\alpha_0(x, v)$. Projecting ψ_0 into the tangent space simply centers it, so that it has zero expectation. The projection of ψ_0 into the tangent space is $\psi_0^\dagger = \varepsilon\alpha_0(x, v)$. For any score I have that $\mathrm{E}_{P_0}\left[S_0|X_i, V_i\right] = 0$. Therefore, $\psi_0 - \psi_0^\dagger$ is orthogonal to the tangent space:

$$\begin{aligned}\mathrm{E}_{P_0}\left[(\psi_0 - \psi_0^\dagger)S_0\right] &= \mathrm{E}_{P_0}\left[F_0(\theta_0(X_i), V_i)\alpha_0(X_i, V_i)S_0\right] = \\ &= \mathrm{E}_{P_0}\left[F_0(\theta_0(X_i), V_i)\alpha_0(X_i, V_i)\,\mathrm{E}_{P_0}\left[S_0|X_i, V_i\right]\right] = 0.\end{aligned}$$

Hence, I conclude that the semiparametric efficiency bound for γ_0^* in the model $\mathcal{P}'$ with additional restrictions is $\mathrm{E}_{P_0}\left[\left(\psi_0^\dagger\right)^2\right] = \mathrm{E}_{P_0}\left[\varepsilon_i^2 \alpha_0(X_i, V_i)^2\right]$. By Lemma 5.1, the bound under the assumptions in the statement of the theorem, denoted $\mathbf{V}_{\gamma_0^*}$, will be no smaller that this amount. That is, by the Law of Iterated Expectations,

$$\mathbf{V}_{\gamma_0^*} \geq \mathrm{E}_{P_0}\left[\varepsilon_i^2 \alpha_0(X_i, V_i)^2\right] = \mathrm{E}_{P_0}\left[\alpha_0(X_i, V_i)^2 \, \mathrm{E}_{P_0}\left[(Y_i - F_0(\theta_0(X_i), V_i))^2 | X_i, V_i\right]\right] = \\ = \mathrm{E}_{P_0}\left[\alpha_0(X_i, V_i)^2 \sigma_Y^2(X_i, V_i)\right].$$

Therefore, if the RHS is infinite, the semiparametric information bound for the CASF in the model $\mathcal{P}$ (where no additional restrictions have been added) is infinite, too. That is, in this case, the CASF is irregularly identified. $\square$

Proof of Lemma 5.2.
Under the above assumptions, the conditional distribution of T_i given $V_i = v$ is

$$F_0^{t|v}(t, v) \equiv P_0(T_i \leq t | V_i = v) = P_0(h\tilde{\pi}_0(Z_i) + V_i \leq t | V_i = v) = \\ = P_0(\tilde{\pi}_0(Z_i) \leq (t - v)/h | V_i = v) = F_0^{\pi}((t - v)/h),$$

where F_0^π is the marginal distribution of $\tilde{\pi}_0(Z_i)$. Hence, taking derivatives one gets $f_0^{t|v}(t, v) = h^{-1} f_0^\pi((t - v)/h)$. That is, the conditional distribution has the shape of an approximation of the Dirac delta.

Since the joint density of (T_i, V_i) can be written as $f_0^{tv} = f_0^{t|v} f_0^v$:

$$\mathrm{E}_{P_0}\left[\alpha_0(T_i, V_i)^2\right] = \int_{\mathbb{R}^2} \frac{f^{t*}(t)^2 f_0^v(v)}{f_0^{t|v}(t, v)} dm(t, v) = \int_{\mathbb{R}^2} \frac{f^{t*}(t)^2 f_0^v(v)}{h^{-1} f_0^\pi((t - v)/h)} dm(t, v), \tag{A.1}$$

where m is the Lebesgue measure. Now, the key fact is that, for small enough h, the density in the denominator can be made arbitrarily small while the numerator is still large.

Specifically, since f_0^π is integrable, $f_0^\pi(\pi) = o(|\pi|^{-1})$ as $|\pi| \to \infty$. Therefore, as long as $t \neq v$, $f_0^\pi((t - v)/h) = o(h)$ as $h \to 0$. That is, for any ε, there exists a $\bar{h}$ such that $h^{-1} f_0^\pi((t - v)/h) < \varepsilon$ if $h < \bar{h}$.

Take also an arbitrary lower bound $\underline{f}^t > 0$ for the counterfactual density f^{t*}. Define $\bar{t} \equiv \sup\{|t| : f^{t*}(t) \geq \underline{f}^t\}$. This supremum is finite since the tails of f^{t*} must converge to zero. One can equivalently define a $\bar{v}$ such that $|v| \leq \bar{v}$ implies $f_0^v(v) \geq \underline{f}^v > 0$. With this, define the set

$$A \equiv \{(t, v) \in \mathbb{R}^2 : |t| \leq \bar{t}, v \leq \bar{v}, t \neq v\},$$

which has positive Lebersgue measure. Then, following (A.1),

$$\mathrm{E}_{P_0}\left[\alpha_0(T_i, V_i)^2\right] \geq \int_A \frac{f^{t*}(t)^2 f_0^v(v)}{h^{-1} f_0^\pi((t - v)/h)} dm(t, v) \geq \frac{\left(\underline{f}^t\right)^2 \underline{f}^v m(A)}{\varepsilon}.$$

Since ε can be taken arbitrarily small, this concludes the proof. $\square$

Proof of Corolary 5.1.
If $\sigma_Y^2(x,v) \geq C > 0$, then

$$\mathrm{E}_{P_0}\left[\alpha_0(X_i,V_i)^2\sigma_Y^2(X_i,V_i)\right] \geq C\,\mathrm{E}_{P_0}\left[\alpha_0(X_i,V_i)^2\right] \to \infty$$

as $h \to 0$. By Theorem 5.1, the CASF will then be irregularly identified. $\square$

Proof of Proposition 6.1.
Since $\sigma_Y^2(t,v) = F_0(\theta_{00}+\theta_{10}t,v)(1-F_0(\theta_{00}+\theta_{10}t,v))$, I am looking for conditions under which

$$\begin{aligned} &\mathrm{E}_{P_0}\left[\alpha_0(T_i,V_i)^2F_0(\theta_{00}+\theta_{10}T_i,V_i)(1-F_0(\theta_{00}+\theta_{10}T_i,V_i))\right] = \\ &\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\alpha_0(t,v)^2F_0(\theta_{00}+\theta_{10}t,v)(1-F_0(\theta_{00}+\theta_{10}t,v))f_0^{tv}(t,v)dvdt \end{aligned} \tag{A.2}$$

is finite. To simplify the computations, I define

$$r \equiv 1-R^2 = \frac{\sigma_V^2}{\sigma_T^2},$$

the share of the endogenous regressor's variance unexplained by the instrument. First, for a normal P_0 and the counterfactual distribution $T_i =^d N(0,\sigma_{T^*}^2)$, I have that:

$$\alpha_0(t,v)^2f_0^{tv}(t,v) = \frac{\sqrt{1-r}\sigma_T}{2\pi\sigma_{T^*}^2\sigma_V}\underbrace{e^{-\frac{1}{2(1-r)}\left[\left(\frac{2(1-r)}{\sigma_{T^*}^2}-\frac{1}{\sigma_T^2}\right)t^2+2\frac{tv}{\sigma_T^2}+\frac{1-2r}{\sigma_V^2}v^2\right]}}_{\equiv E_1(t,v)}. \tag{A.3}$$

Second, under the joint normality assumption:

$$F_0(u,v) = \frac{1}{\sqrt{2\pi}\sigma_{U|V}}\int_{-\infty}^{u}e^{-\frac{1}{2\sigma_{U|V}^2}\left(s-\rho_{U|V}v\right)^2}ds.$$

To bound the above distribution, I introduce here inequalities 7.11 and 7.17 in Abramowitz and Stegun (1972): for $x \geq 0$,

$$\frac{e^{-x^2}}{x+\sqrt{x^2+2}} \leq \int_x^\infty e^{-u^2}du \leq \frac{e^{-x^2}}{x+\sqrt{x^2+4/\pi}}.$$

These inequalities allow me to bound both $F_0(\theta_{00}+\theta_{10}t,v)$ and $1-F_0(\theta_{00}+\theta_{10}t,v)$. To ease notation, call $\xi(t,v) \equiv \theta_{00}+\theta_{10}t-\rho_{U|V}v$. Moreover, define

$$\kappa_j(x) \equiv \frac{1}{\sqrt{\pi}}\frac{1}{\frac{x}{\sqrt{2}\sigma_{U|V}}+\sqrt{\frac{x^2}{2\sigma_{U|V}^2}+c_j}}, \text{ where } c_j = \begin{cases} 2 & \text{if } j=1 \\ 4/\pi & \text{if } j=2 \end{cases}.$$

Then, if $\xi(t,v) \leq 0$:

$$e^{-\frac{\xi(t,v)^2}{2\sigma^2_{U|V}}}\kappa_1(-\xi(t,v)) \leq F(\theta_{00}+\theta_{10}t,v) \leq e^{-\frac{\xi(t,v)^2}{2\sigma^2_{U|V}}}\kappa_2(-\xi(t,v)), \tag{A.4}$$

$$\frac{1}{2} \leq 1 - F(\theta_{00}+\theta_{10}t,v) \leq 1; \tag{A.5}$$

and, if $\xi(t,v) \geq 0$:

$$\frac{1}{2} \leq F(\theta_{00}+\theta_{10}t,v) \leq 1, \tag{A.6}$$

$$e^{-\frac{\xi(t,v)^2}{2\sigma^2_{U|V}}}\kappa_1(\xi(t,v)) \leq 1 - F(\theta_{00}+\theta_{10}t,v) \leq e^{-\frac{\xi(t,v)^2}{2\sigma^2_{U|V}}}\kappa_2(\xi(t,v)). \tag{A.7}$$

I now proceed as follows. Recall that E_1 was defined in equation (A.3). Then, I define:

$$\begin{aligned} I \equiv \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E_1(t,v)F_0(\theta_{00}+\theta_{10}t,v)(1-F_0(\theta_{00}+\theta_{10}t,v))dvdt = \\ \iint_{\{\xi(t,v)\leq 0\}} E_1(t,v)F_0(\theta_{00}+\theta_{10}t,v)(1-F_0(\theta_{00}+\theta_{10}t,v))dvdt + \\ \iint_{\{\xi(t,v)> 0\}} E_1(t,v)F_0(\theta_{00}+\theta_{10}t,v)(1-F_0(\theta_{00}+\theta_{10}t,v))dvdt = \\ I_1 + I_2. \end{aligned}$$

Clearly, the integral in (A.2) is finite if and only if I is finite. Moreover, using inequalities (A.4) and (A.5), I can bound the first of the integrals in the last row of the above display (I_1). Equivalently, the second one (I_2) is bounded using inequalities (A.6) and (A.7):

$$\frac{1}{2}\iint_{\{\xi(t,v)\leq 0\}} \kappa_1(-\xi(t,v))E_1(t,v)e^{-\frac{\xi(t,v)^2}{2\sigma^2_{U|V}}}dvdt \leq I_1 \leq \iint_{\{\xi(t,v)\leq 0\}} \kappa_2(-\xi(t,v))E_1(t,v)e^{-\frac{\xi(t,v)^2}{2\sigma^2_{U|V}}}dvdt,$$

$$\frac{1}{2}\iint_{\{\xi(t,v)> 0\}} \kappa_1(\xi(t,v))E_1(t,v)e^{-\frac{\xi(t,v)^2}{2\sigma^2_{U|V}}}dvdt \leq I_2 \leq \iint_{\{\xi(t,v)> 0\}} \kappa_2(\xi(t,v))E_1(t,v)e^{-\frac{\xi(t,v)^2}{2\sigma^2_{U|V}}}dvdt.$$

Therefore, I can conclude that:

$$\frac{1}{2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \kappa_1(|\xi(t,v)|)E_1(t,v)e^{-\frac{\xi(t,v)^2}{2\sigma^2_{U|V}}}dvdt \leq I \leq \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \kappa_2(|\xi(t,v)|)E_1(t,v)e^{-\frac{\xi(t,v)^2}{2\sigma^2_{U|V}}}dvdt. \tag{A.8}$$

Now, to analyze the behavior of the above bounds, define the functions

$$\eta_j(x) \equiv \kappa_j(|x|)e^{-\frac{\theta_{00}(2x-\theta_{00})}{2\sigma^2_{U|V}}} = \frac{1}{\sqrt{\pi}}\frac{e^{-\frac{\theta_{00}(2x-\theta_{00})}{2\sigma^2_{U|V}}}}{\frac{|x|}{\sqrt{2}\sigma_{U|V}} + \sqrt{\frac{x^2}{2\sigma^2_{U|V}} + c_j}} \tag{A.9}$$

and the matrix

$$M \equiv \begin{pmatrix} \frac{2(1-r)}{\sigma_{T^*}^2} - \frac{1}{\sigma_T^2} + \frac{1-r}{\sigma_{U|V}^2}\theta_{10}^2 & \frac{1}{\sigma_T^2} - \frac{1-r}{\sigma_{U|V}^2}\theta_{10}\rho_{U|V} \\ \frac{1}{\sigma_T^2} - \frac{1-r}{\sigma_{U|V}^2}\theta_{10}\rho_{U|V} & \frac{1-2r}{\sigma_V^2} + \frac{1-r}{\sigma_{U|V}^2}\rho_{U|V}^2 \end{pmatrix}.$$

With these definitions, I can rewrite equation (A.8) as:

$$\frac{1}{2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\eta_1(\xi(t,v))e^{-\frac{1}{2(1-r)}(t,v)M(t,v)'}dvdt \le I \le \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\eta_2(\xi(t,v))e^{-\frac{1}{2(1-r)}(t,v)M(t,v)'}dvdt.$$

In the above equation, the dominant term is the exponential of the quadratic form $(t,v)M(t,v)'$. This is confirmed by the following lemma:

Lemma A.1. *For η_j, $j \in \{1,2\}$, as defined in equation* (A.9)*:*

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\eta_j(\xi(t,v))e^{-\frac{1}{2(1-r)}(t,v)M(t,v)'}dvdt < \infty \Leftrightarrow M \textit{ is positive definite.}$$

Proof. Let $\mathcal{I}_j \equiv \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\eta_j(\xi(t,v))e^{-\frac{1}{2(1-r)}(t,v)M(t,v)'}dvdt$. Suppose that M is not positive definite. I define region $\mathcal{R}$ by the points satisfying $(t,v)M(t,v)' \le 1$. Then, since η_j are positive:

$$\mathcal{I}_j \ge e^{-\frac{1}{2(1-r)}}\iint_{\mathcal{R}}\eta_j(\xi(t,v))d(t,v).$$

In the case M is negative definite or semidefinite, $\mathcal{R} = \mathbb{R}^2$. If M is indefinite, $\mathcal{R}$ is the region between the two branches of an hyperbola. If M is positive semidefinite, $\mathcal{R}$ is the region between two parallel lines.

Consider now the half plane given by $\{\xi(t,v) \le -K\}$. Since M is not positive definite, $\mathcal{R}_K \equiv \mathcal{R} \cap \{\xi(t,v) \le -K\}$ has infinite Lebesgue measure in $\mathbb{R}^2$ for any $K > 0$. Moreover,

$$\lim_{x\to-\infty}\eta_j(x) = \infty.$$

Thus, $\forall K > 0$, there exists $L_j > 0$ such that $x < -K \Rightarrow \eta_j(x) > L_j$. Hence, I can conclude that

$$\mathcal{I}_j \ge e^{-\frac{1}{2(1-r)}}\iint_{\mathcal{R}_K}\eta_j(\xi(t,v))d(t,v) \ge e^{-\frac{1}{2(1-r)}}L_j\iint_{\mathcal{R}_K}d(t,v) = \infty.$$

Assume now that M is positive definite. Since $\kappa_j(|x|) \le \frac{1}{\sqrt{\pi c_j}}$,

$$\begin{aligned}\mathcal{I}_j &\le \frac{1}{\sqrt{\pi c_j}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{-\frac{\theta_{00}(2\xi(t,v)-\theta_{00})}{2\sigma_{U|V}^2}-\frac{1}{2(1-r)}(t,v)M(t,v)'}dvdt = \\ &= C\int_{\mathbb{R}^2}e^{(t,v)m'-(t,v)\tilde{M}(t,v)'}d(t,v),\end{aligned}$$

for a constant $C \equiv \frac{1}{\sqrt{\pi c_j}}e^{-\frac{\theta_{00}^2}{2\sigma_{U|V}^2}}$, the constant vector $m \equiv \frac{\theta_{00}}{\sigma_{U|V}}(-\theta_{10}, \rho_{U|V})$ and the rescaled matrix $\tilde{M} \equiv \frac{1}{2(1-r)}M$.

Since $1-r > 0$, $\tilde{M}$ is positive definite. Therefore, there exists a diagonal matrix $D =$

diag(λ_1, λ_2), with $\lambda_1, \lambda_2 > 0$, and an orthogonal matrix P such that $\tilde{M} = PDP'$. I can now perform a change of variables $(s_1, s_2) \equiv (t, v)P$, so that (being $(\tilde{m}_1, \tilde{m}_2) \equiv mP$):

$$\mathcal{I}_j \leq C \left(\int_{-\infty}^{\infty} e^{\tilde{m}_1 s_1 - \lambda_1 s_1^2} ds_1 \right) \left(\int_{-\infty}^{\infty} e^{\tilde{m}_2 s_2 - \lambda_2 s_2^2} ds_2 \right).$$

The RHS of the above equation is finite since both eigenvalues are positive. $\square$

I now proceed to check under which conditions is M positive definite. Recall that $\rho_{UV} \equiv \frac{\sigma_{UV}}{\sigma_U \sigma_V}$ and $\Theta \equiv \frac{\sigma_T}{\sigma_U}\theta_{10}$. M is positive definite if and only if the following two conditions are satisfied:

$$\frac{1-2r}{\sigma_V^2} + \frac{1-r}{\sigma_{U|V}^2}\rho_{U|V}^2 > 0, \tag{A.10}$$

$$\det(M) > 0. \tag{A.11}$$

Condition (A.10) is equivalent to

$$1 - (2 - \rho_{UV}^2)r > 0 \Leftrightarrow r < \frac{1}{2 - \rho_{UV}^2}.$$

Condition (A.11) is equivalent to

$$\begin{aligned}
&\frac{2(1-r)(1-2r)}{\sigma_V^2}\frac{1}{\sigma_{T^*}^2} + \frac{2(1-r)^2\rho_{U|V}^2}{\sigma_{U|V}^2}\frac{1}{\sigma_{T^*}^2} - \frac{1-2r}{\sigma_V^2\sigma_T^2} \\
&- \frac{(1-r)\rho_{U|V}^2}{\sigma_{U|V}^2\sigma_T^2} + \frac{(1-r)(1-2r)}{\sigma_{U|V}^2\sigma_V^2}\theta_{10}^2 + \frac{2(1-r)\rho_{U|V}}{\sigma_{U|V}^2\sigma_T^2}\theta_{10} > \frac{1}{\sigma_T^4} \\
&\Updownarrow \\
&\frac{\sigma_T^2}{\sigma_{T^*}^2}\frac{2(1-r)(1-2r)}{r} + \frac{\sigma_T^2}{\sigma_{T^*}^2}\frac{\rho_{UV}^2}{1-\rho_{UV}^2}\frac{2(1-r)^2}{r} - \frac{1-2r}{r} \\
&- \frac{\rho_{UV}^2}{1-\rho_{UV}^2}\frac{1-r}{r} + \frac{1}{1-\rho_{UV}^2}\frac{(1-r)(1-2r)}{r}\Theta^2 + \frac{\rho_{UV}}{1-\rho_{UV}^2}\frac{2(1-r)\sqrt{r}}{r}\Theta > 1 \\
&\Updownarrow \\
&\left[1 - \Theta^2 + 2\sqrt{r}\Theta\left(\sqrt{r}\Theta - \rho_{UV}\right)\right]\frac{\sigma_{T^*}^2}{\sigma_T^2} < 2\left[1 - (2-\rho_{UV}^2)r\right].
\end{aligned}$$

Note that the condition in Proposition 6.1, given in equation (6.3), is obtained by replacing $r = 1 - R^2$ in the final row of the above display. To conclude the proof, it is left to show that condition (A.11) implies condition (A.10). In this case, it is enough to check the former. The following lemma takes care of it:

Lemma A.2. *For $r \in (0,1)$, $\rho_{UV} \in (-1,1)$ and $\Theta \in \mathbb{R}$:*

$$\left[1 - \Theta^2 + 2\sqrt{r}\Theta\left(\sqrt{r}\Theta - \rho_{UV}\right)\right]\frac{\sigma_{T^*}^2}{\sigma_T^2} < 2\left[1 - (2-\rho_{UV}^2)r\right] \Rightarrow 1 - (2-\rho_{UV}^2)r > 0.$$

Proof. Define $H(r, \rho_{UV}, \Theta) \equiv 1 - \Theta^2 + 2\sqrt{r}\Theta\left(\sqrt{r}\Theta - \rho_{UV}\right)$ and $K(r, \rho_{UV}) \equiv 1 - (2 - \rho_{UV}^2)r$. The roots of H, if they exists, are given by:

$$\sqrt{r_1} = \frac{\rho_{UV} - \sqrt{\rho_{UV}^2 - 2(1 - \Theta^2)}}{2\Theta} \text{ and } \sqrt{r_2} = \frac{\rho_{UV} + \sqrt{\rho_{UV}^2 - 2(1 - \Theta^2)}}{2\Theta}. \tag{A.12}$$

Furthermore, $K(r, \rho_{UV}) > 0$ if $r < r^*$, where $r^* \equiv \frac{1}{2-\rho_{UV}^2}$.

I assume that $H(r, \rho_{UV}, \Theta)\frac{\sigma_{T^*}^2}{\sigma_T^2} < K(r, \rho_{UV})$ holds. In this case, since the variance ratio is always positive:

$$H(r, \rho_{UV}, \Theta) \geq 0 \Rightarrow K(r, \rho_{UV}) > 0.$$

Thus, the $H(r, \rho_{UV}, \Theta) < 0$ cases are left to study. As $H(r, -\rho_{UV}, -\Theta) = H(r, \rho_{UV}, \Theta)$ and $K(r, -\rho_{UV}) = K(r, \rho_{UV})$, without loss of generality I focus on $\Theta \geq 0$. I analyze under which cases can the LHS be negative:

- $0 \leq \Theta < 1$:
  - When $\rho_{UV} \leq 0$, $H(r, \rho_{UV}, \Theta) \geq 0$ for any $r \in (0, 1)$.
  - When $\rho_{UV} > 0$, there is a further condition on the size of Θ:
    - If $\Theta^2 \leq 1 - \frac{\rho_{UV}^2}{2}$, then $H(r, \rho_{UV}, \Theta) \geq 0$ for any $r \in (0, 1)$.
    - If $\Theta^2 > 1 - \frac{\rho_{UV}^2}{2}$, then $H(r, \rho_{UV}, \Theta)$ is negative for $r \in (r_1, r_2)$.
- $\Theta \geq 1$: $H(r, \rho_{UV}, \Theta)$ is negative for $r < r_2$.

To sum up, to conclude the proof I need to show that $r_2 \leq r^*$, so that $K(r, \rho_{UV}) > 0$ for any r such that $H(r, \rho_{UV}, \Theta) < 0$. First, from the discussion above, if H is negative, it must be that $\Theta^2 > 1 - \frac{\rho_{UV}^2}{2} = \frac{2-\rho_{UV}^2}{2}$. Second, r_2 as a function of Θ attains a maximum when

$$\rho_{UV}^2 - (1 - \Theta) + \rho_{UV}\sqrt{\rho_{UV}^2 - 2(1 - \Theta^2)} = 1.$$

Therefore,

$$r_2 = \frac{\rho_{UV}^2 - (1 - \Theta) + \rho_{UV}\sqrt{\rho_{UV}^2 - 2(1 - \Theta^2)}}{2\Theta^2} \leq \frac{1}{2\Theta^2} \leq \frac{1}{2 - \rho_{UV}^2} = r^*.$$

□

□

Proof of Corollary 6.1.
The fact that the CASF is irregularly identified if condition (6.3) is not satisfied follows from Theorem 5.1. The theorem can be applied since P_0 as in the assumptions of Proposition 6.1 satisfies Assumptions 3.1, 4.1, and 4.2.

To show that the bound on the ratio of variances has the stated shape, note that if $H(R^2, \rho_{UV}, \Theta) > 0$, then

$$\frac{\sigma^2_{T^*}}{\sigma^2_T} < \frac{1-(2-\rho^2_{UV})(1-R^2)}{H(R^2, \rho_{UV}, \Theta)}$$

for the CASF to be regularly identified. However, the LHS of condition (6.3), $H(R^2, \rho_{UV}, \Theta)$, may be negative. Nevertheless, in that case the RHS is positive (see the proof of Lemma A.2):

$$H(R^2, \rho_{UV}, \Theta) < 0 \Rightarrow 1-(2-\rho^2_{UV})(1-R^2) \geq 0.$$

Thus, for those cases condition (6.3) is satisfied for any value of $\sigma^2_{T^*}/\sigma^2_T$.

□

Proof of Corollary 6.2.
Irregular identification of the CASF follows by Theorem 5.1, as it did in the proof of Corollary 6.1. To obtain the lower bound, I will again work with $r \equiv 1-R^2$, as I did in the proof of Proposition 6.1. I define, as in the proof of Lemma A.2, $H(r, \rho_{UV}, \Theta) \equiv 1-\Theta^2+2\sqrt{r}\Theta\left(\sqrt{r}\Theta-\rho_{UV}\right)$ and $K(r, \rho_{UV}) \equiv 1-(2-\rho^2_{UV})r$. The roots of H (r_1 and r_2), if they exists, are given by equation (A.12). Let $r^* \equiv \frac{1}{2-\rho^2_{UV}}$.

If $H(r, \rho_{UV}, \Theta) > 0$, the LHS of condition (6.3) is positive. Therefore, the CASF is not regularly identified if the RHS, given by $K(r, \rho_{UV})$, is non-positive. This happens when $r \geq r^*$.

To conclude, I need to argue that $H(r^*, \rho_{UV}, \Theta) > 0$ for any (ρ_{UV}, Θ). If this was not the case, it may be that both sides of condition (6.3) are negative. The above claim is true by the arguments in the proof of Lemma A.2. Indeed, $H(r, \rho_{UV}, \Theta) > 0$ if $r > r_2$, and $r^* > r_2$..

Therefore, the CASF is irregularly identified if $r \geq r^*$. This implies that the CASF is irregularly identified if

$$1-R^2 \geq \frac{1}{2-\rho^2_{UV}} \Leftrightarrow R^2 \leq 1-\frac{1}{2-\rho^2_{UV}} \equiv \underline{R}^2.$$

□

Proof of Proposition 6.2.
For a normal P_0 and the counterfactual distribution $T_i =^d N(\mu_{T^*}, \sigma^2_T)$, being $r \equiv 1-R^2$:

$$\alpha_0(t,v)^2 f_0^{tv}(t,v) = \frac{\sqrt{1-r}}{2\pi\sigma_T\sigma_V} e^{\frac{2\mu_{T^*}t-\mu^2_{T^*}}{\sigma^2_T}} \underbrace{e^{-\frac{1}{2(1-r)}\left(\frac{1-2r}{\sigma^2_T}t^2+2\frac{tv}{\sigma^2_T}+\frac{1-2r}{\sigma^2_V}v^2\right)}}_{\equiv E_1(t,v)}.$$

The E_1 defined in the above equation corresponds to the E_1 defined in the proof of Proposition 6.1, for the particular case where $\sigma^2_{T^*} = \sigma^2_T$. I can, therefore, follow the reasoning in the proof of Proposition 6.1. The result given in Lemma A.1 does not rely on the specific coefficients of the linear term in the exponential (these are absorbed in η_j). Thus, it does hold for this case. I conclude that the condition obtained in the proof of Proposition 6.1 is also valid for the current case, evaluating it at $\sigma^2_{T^*}/\sigma^2_T = 1$.

$\square$