(2016).

In order to ensure that such estimators are well-defined and stable, it is important to assess the overlap between the distribution of the propensity score among treatment and control groups, i.e. to check whether the propensity score is bounded away from zero and one, and if the support of the propensity score in both groups are nearly the same, see e.g. Heckman, Ichimura, Smith and Todd (1998), Smith and Todd (2005), Crump et al. (2009) and Khan and Tamer (2010). Following Heckman, Ichimura, Smith and Todd (1998), it is now routine to compare kernel density estimates of the propensity score among treated and control samples to determine the common support region. In cases where there is strong overlap one proceeds as described above, otherwise, one usually considers trimmed samples, see e.g. Crump et al. (2009) and Sasaki and Ura (2018).

Although kernel density estimators are popular, they involve choosing tuning parameters such as bandwidths and often suffer from boundary bias. Of course such inconveniences can be easily avoided if one focuses on CDFs instead of densities. In the following we show that propensity score overlap implies a particular set of restrictions between the CDFs of treated and control groups, and that these restrictions can form the basis for testing the correct specification of propensity score models.

Assume that the propensity score $p(X)$ has a density with respect to a dominating measure, and that the density is bounded away from zero and infinity uniformly over its support. The following lemma builds on Shaikh et al. (2009) and formalizes the above discussion.

Lemma 1 *Let* $\alpha = \mathbb{P}(D=0)/\mathbb{P}(D=1)$ *and assume that* $0 < \mathbb{P}(D=1) < 1$*. If* $0 < p(X) < 1$ *a.s., then*

$$\mathbb{E}\left[1\{p(X) \le u\} | D=1\right] = \alpha \, \mathbb{E}\left[\frac{p(X)}{1-p(X)} 1\{p(X) \le u\} | D=0\right], \ \forall u \in [0,1]. \tag{2.1}$$

Furthermore, (2.1) holds if and only if

$$\mathbb{E}\left[(D - p(X)) 1\{p(X) \le u\}\right] = 0, \ \forall u \in [0,1]. \tag{2.2}$$

Lemma 1 implies that, when the propensity score is correctly specified, one can expect that the sample analogue of (2.1) should hold. Thus, (2.1) provides a graphical diagnostic tool for propensity score misspecification; see Lemma 3.2 of Słoczyński and Wooldridge (2018) for a result related to (2.1). Perhaps more importantly, note that (2.2) provides an infinite number of simple unconditional moment restrictions that can be used to formally test whether