An endogeneity correction based on a nonparametric control function approach

Jörg Breitung and Dominik Wied
University of Cologne

Abstract
This paper considers a linear regression model with an endogenous regressor which is not normally distributed. It is shown that the corresponding coefficient can be consistently estimated without external instruments by adding a rank-based transformation of the regressor to the model and performing standard OLS estimation. In contrast to other approaches, our nonparametric control function approach does not rely on a conformably specified copula. Furthermore, the approach allows for the presence of additional exogenous regressors which may be (linearly) correlated with the endogenous regressor(s). Consistency and further asymptotic properties of the estimator are considered and the estimator is compared with copula based approaches by means of Monte Carlo simulations. An empirical application on wage data of the US current population survey demonstrates the usefulness of our method.

Keywords: IV regression, generated regressors, wage equations
1. INTRODUCTION

Employing instrumental variables is a classical econometric tool for identifying causal effects if no randomized controlled trial or suitable observed control variables are available (see e.g. Angrist and Pischke, 2008). Instruments allow for constructing quasi-experiments and yield consistent parameter estimators for endogenous regressors as long as they are uncorrelated with the error term. In empirical practice, however, a common problem is the choice of suitable instruments. In many situations it is not easy to find appropriate instrumental variables and sometimes there are many possible candidates so that it is not clear how to make the choice.

The present paper studies a new approach for constructing consistent estimators in a regression with endogenous regressors without using (external) instruments. As in Park and Gupta (2012) we assume that the distribution of the endogenous regressor is different from that of the error term with the leading case that the regressor is not normally distributed, while the error term is. In contrast to the copula-based approach in earlier work, our endogeneity correction relies on a nonparametric control function approach. We just assume that the endogeneity is due to a normally distributed component in the regression error that enters the endogenous regressor in a nonlinear way. Our approach does not need to assume that the nonlinear relationship between the endogenous regressor and the error term is known. Rather we just assume that it is monotonously increasing or decreasing. Hence our endogeneity adjustment is basically nonparametric, although the underlying regression model is fully parametric.

The idea goes back at least to Park and Gupta (2012) who propose a related copula-based approach. In their estimation method, the joint distribution of the endogenous regressor and the error is represented by a copula. The parameters are estimated by maximum likelihood, where the likelihood function is based on the estimated joint distribution. There are two potential drawbacks in their approach. On the one hand, they do not study the asymptotic
properties of their estimator and hence it is not clear under which conditions their approach is consistent. On the other hand, they (implicitly) assume that any additional exogenous regressor is independent of the endogenous regressor. These drawbacks are partly addressed by Haschka (2021) and Yang, Qian, and Xie (2022) who allow for a particular form of dependence between the exogenous and endogenous regressors. However, their proof for consistency assumes that the distribution of the endogenous regressor is known which is typically not the case in empirical practice. Furthermore they do not study other asymptotic properties such as the convergence rate and asymptotic normality of their estimator. Haschka (2022) provides an extension of the idea to linear panel models based on maximum likelihood, but does not study the asymptotic properties.

Our paper goes beyond the existing work by proposing an estimator, for which we do not have to specify a suitable copula. Furthermore, it allows for a correlation between the endogenous regressor and other exogenous regressors and we rigorously study the asymptotic properties of our estimator. The estimator is calculated in two steps: First, we remove the (linear) dependence among the exogenous and endogenous regressors by computing the residuals from an auxiliary regression. Second, the endogenous part of the error is estimated by applying the quantile function of the normal distribution on the ranks of the residuals from the first step. In contrast, Yang et al. (2022) (2sCOPE) calculate residuals from a regression of the transformed regressors, thereby assuming a particular nonlinear dependence among the exogenous and endogenous regressors.

Due to the two-step estimation, it is not straightforward to derive asymptotic results for our estimator. A particular problem is that both the residuals and the ranks are dependent within each other so that no simple law of large number or central limit theorem can be applied. We rely on results for stochastic integrals in order to analyze the asymptotic properties. In particular, we derive consistency of the structural parameters.

The plan of our paper is as follows. In Section 2 we first develop our nonparametric control
function approach, before the empirical endogeneity correction is introduced in Section 3. A link to IV estimation is provided in Section 4 and asymptotic properties are derived in Section 5. Next we study the small sample properties by means of Monte Carlo simulation in Section 6 and provide an empirical application to wage data (Section 7). This application is well suited for our approach because the potentially endogenous regressor (years of education) is apparently not normal and because there is an ongoing debate in the empirical literature about the choice of appropriate instruments.

2. THE NONPARAMETRIC CONTROL FUNCTION APPROACH

Consider a linear regression

\[ y_i = x_i' \beta + z_i \gamma + u_i \]  

(2.1)

where we assume that \( x_i \) is a \( k \times 1 \) vector of exogenous variables and \( z_i \) is an endogenous variable that is correlated with the error \( u_i \). We focus on a single endogenous regressor just for notational convenience. The treatment of more endogenous regressors turns out to be straightforward by applying the endogeneity correction to each endogenous regressor separately. Assume further that the relationship between the endogenous regressor \( z_i \) and the error term \( u_i \) can be represented as

\[ u_i = f(z_i) + \varepsilon_i \]  

(2.2)

where \( \varepsilon_i \) is independent of \( z_i \) and \( x_i \). Since we assume that \( x_i \) is exogenous we have \( E(x_iu_i) = 0 \) and, therefore, \( E[f(z_i)x_i] = 0. \)

The traditional (IV) approach assumes that \( f(z_i) \) is a linear function and there exists an \( m \times 1 \) vector of instrumental variables \( v_i \) such that \( E(v_iz_i) \neq 0 \) and \( E(v_iu_i) = 0. \) The relationship between the endogenous regressor and the instrumental variables is given by the (first stage)
regression

\[ z_i = x_i'\pi_1 + v_i'\pi_2 + \eta_i . \] (2.3)

Since \( x_i \) and \( v_i \) are exogenous, they do not enter the linear representation for \( u_i \) and, thus, the endogeneity is represented by the decomposition

\[ u_i = \rho \eta_i + \varepsilon_i . \] (2.4)

Accordingly, the control function \( f(z_i) = \rho(z_i - x_i'\pi_1 - v_i'\pi_2) = \rho \eta_i \) in (2.2) is linear and requires an external instrument \( v_i \). Inserting (2.4) in (2.1) yields

\[ y_i = x_i'\beta + z_i\gamma + \rho \eta_i + \varepsilon_i . \] (2.5)

In this regression equation the additional variable \( \eta_i \) controls for the endogeneity of \( z_i \), as the remaining error \( \varepsilon_i \) is uncorrelated with all regressors. In practice \( \eta_i \) is replaced by the residual of the first stage regression (2.3) denoted by \( \hat{\eta}_i \). It is not difficult to show that the OLS regression (2.5) with \( \hat{\eta}_i \) instead of \( \eta_i \) is equivalent to the usual IV (or two-stage LS) estimator, cf. Wooldridge (2015).

The crucial identifying assumption in the linear model is that the instrument \( v_i \) does not enter the structural model (2.1), which is equivalent to the assumption that \( v_i \) is uncorrelated with \( u_i \). Park and Gupta (2012) adopt a different identifying assumption and assume that \( f(\cdot) \) in (2.2) is a nonlinear function. Let us assume for a moment that this function is known. It is important to note that \( f(z_i) \) obeys the restriction \( \mathbb{E}[x_i f(z_i)] = 0 \) as we assume that \( x_i \) is exogenous with respect to \( u_i \). To this end we assume that the error and the endogenous variable can be represented as

\[ u_i = f(e_i) + \varepsilon_i \] (2.6)

and \( z_i = \delta'x_i + e_i \) (2.7)
where \( e_i \) represents the component in \( z_i \) that is uncorrelated with \( x_i \), whereas \( \delta'x_i \) represents the (linear) dependence between \( z_i \) and \( x_i \). The crucial identifying assumptions are that (i) \( f(\cdot) \) is a nonlinear monotonous function and (ii) the error term (and therefore both components \( f(e_i) \) and \( \varepsilon_i \)) are normally distributed with \( \mathbb{E}(u_i) = 0 \) and \( \mathbb{E}(u_i^2) = \sigma_u^2 \). These assumptions are sufficient to identify the nonlinear function \( f(\cdot) \) up to a scale factor. Let \( F_e(\cdot) \) denote the c.d.f. of the endogenous regressor \( e_i \). Then, \( F_e(e_i) \) is a uniformly distributed random variable in the interval \([0, 1] \). Similarly \( \Phi(f(e_i)/\sigma_f) \) is uniformly distributed in the interval \([0, 1] \), where \( \sigma_f^2 = \mathbb{E}[f(e_i)^2] \) and \( \Phi(\cdot) \) denotes the c.d.f. of the standard normal distribution. Since \( F_e(\cdot) \) and \( \Phi(f(\cdot)/\sigma_f) \) are strictly monotonous functions it follows that

\[
F_e(e_i) = \Phi(f(e_i)/\sigma_f) \quad (2.8)
\]

or

\[
\Phi^{-1}[F_e(e_i)] = \frac{1}{\sigma_f} f(e_i) := \eta_i. \quad (2.9)
\]

Note that a similar result was also employed in Park and Gupta (2012). In what follows we employ the empirical c.d.f. of \( e_i \), whereas Park and Gupta (2012) construct a maximum likelihood estimator based on the Gaussian copula for the joint distribution of \( z_i \) and \( u_i \).

3. AN EMPIRICAL ENDOGENEITY CORRECTION

In the previous section it is argued that the endogeneity of the variable \( z_i \) can be controlled for by augmenting the regression by \( \eta_i = f(e_i)/\sigma_f \) such that \( \eta_i \) is standardized to have unit variance. In empirical practice the function \( f(\cdot) \) and the argument \( e_i \) are unknown. If \( f(e_i) \) were known, the OLS regression

\[
y_i = \beta'x_i + \gamma z_i + \rho \eta_i + \varepsilon_i .
\]

\(^1\)In Park and Gupta (2012), such a linear dependence is ignored and consistency of their estimator requires that \( f(z_i) \) is independent of \( x_i \).

\(^2\)As noted in Park and Gupta (2012), we may assume any other known distribution.
would yield consistent estimates for the parameters $\beta$ and $\gamma$. The estimation of $\eta_i$ consists of two steps. First we eliminate any (linear) dependence between $z_i$ and $x_i$ by running the regression

$$z_i = \delta' x_i + e_i,$$

(3.1)

where the residual $\hat{e}_i$ serves as an estimator for $e_i$ in (2.7). In a second step we estimate $\eta_i$ by the empirical analog of (2.9)

$$\hat{\eta}_i = \Phi^{-1}[\hat{F}_\varepsilon(\hat{e}_i)],$$

(3.2)

where $\hat{F}_\varepsilon(\cdot)$ denotes the empirical distribution function of $\hat{e}_i$. Note that

$$\hat{F}_\varepsilon(\hat{e}_i) = \frac{1}{n} \text{rank of } (\hat{e}_i).$$

In order to escape the possibility to obtain $\hat{F}_\varepsilon(\hat{e}_i) = 1$ (note that $\Phi^{-1}(a) \to \infty$ as $a \to 1$), we divide by $n + 0.5$. Accordingly our estimate for $\eta_i$ is

$$\hat{\eta}_i = \Phi^{-1}\left[\frac{1}{n + 0.5} \text{rank of } (\hat{e}_i)\right] = \Phi^{-1}\left[\frac{n}{n + 0.5} \hat{F}_\varepsilon(\hat{e}_i)\right].$$

As will be shown in the next section, $\hat{\eta}_i - \eta_i = O_p(n^{-1/2})$ as $n \to \infty$ and thus the correction term is estimated consistently. It should be noted that the estimation error $\hat{\eta}_i - \eta_i$ affects the asymptotic distribution of the estimate of $\beta$ and $\gamma$ (the so-called generated regressor problem). In Section 5 we show that the limiting distribution depends on the unknown nonlinear function $f(\cdot)$.

Although it appears that in empirical practice, standard regression inference ignoring the generated regressor problem already provides a fairly reliable approximation for the actual distribution of the parameter estimates, we propose a bootstrap procedure that is able to

---

\[\text{Note that dividing by } n + 1 \text{ results in } \mathbb{E}(\hat{\eta}_i) = O(n^{-1}). \text{ As long as the regression includes a constant, this does not raise any problems. In a regression without intercept, } \hat{\eta}_i \text{ needs to be demeaned.}\]
reproduce the correct distribution in large samples. To this end we assume that the vector $\mathbf{X}_i = (y_i, x'_i, z_i)'$ is i.i.d. but not necessarily normally distributed. Let $\theta = (\beta', \gamma, \rho)'$ denote the parameter vector of interest. Then our estimator can be represented as $\hat{\theta} = \psi(\mathbf{X}_1, \ldots, \mathbf{X}_n)$, where $\psi(\cdot)$ is some nonlinear function that is invariant to the ordering of the vectors $\mathbf{X}_1, \ldots, \mathbf{X}_n$. Accordingly, the bootstrap distribution of the estimator $\hat{\theta}$ can be obtained by drawing without replacement from the empirical distribution of the vectors $\{\mathbf{X}_1, \ldots, \mathbf{X}_n\}$. If the sample size and the number of bootstrap draws tend to infinity, the bootstrap distribution provides a reliable approximation of the actual distribution of $\hat{\theta}$.

Besides yielding a consistent estimator for the parameter $\gamma$, the approach also allows for testing for endogenous regressors. If $\rho = 0$, the regressor $z_i$ is exogenous, which can be checked with a simple $t$-test for the parameter $\rho$. Such a test procedure is similar to the Durbin-Hausman-Wu test for endogeneity. We will show in Section 5 that the ordinary $t$-test for $\hat{\rho}$ possess a standard normal limiting distribution.

4. THE DUALITY WITH THE IV APPROACH

To simplify the exposition we focus on the model with an endogenous variable only which we write in vector format as

$$y = z\gamma + \hat{\eta}\rho + \tilde{\varepsilon}$$

where $y = (y_1, \ldots, y_n)'$, $z = (z_1, \ldots, z_n)'$, $\hat{\eta} = (\hat{\eta}_1, \ldots, \hat{\eta}_n)'$. The proposed estimator can be represented as

$$\hat{\gamma} = \frac{z'M_{\hat{\eta}}y}{z'M_{\hat{\eta}}z} = \frac{\tilde{v}'y}{\tilde{v}'z} \quad (4.1)$$

where $\tilde{v} = M_{\hat{\eta}}z$ and $M_{\hat{\eta}} = I_n - \hat{\eta}\hat{\eta}'/\hat{\eta}'\hat{\eta}$. This representation of the estimator gives rise to the interpretation of a just-identified IV estimator using the residuals from a regression of $z$ on $\hat{\eta}$ as instrumental variable vector $\tilde{v}$. Accordingly, for the full model we may alternatively
estimate the coefficients $\beta$ and $\gamma$ from an IV regression using $(x_i, \hat{v}_i)$ as instruments with $\hat{v}_i$ as the $i$'th element of the residual vector $\hat{v}$.

This allows us to combine the “internal instrument” $\hat{v}_i$ with possible external instruments resulting in an over-identified IV (or GMM) estimator. Furthermore the approach of Stock and Yogo (2005) may be adapted to test against “weak instruments”. It should be noted, however, that $\hat{v}_i$ is an estimated instrumental variable and its estimation error affects the asymptotic distribution, in general. Although the effect on the asymptotic distribution is typically small, a bootstrap procedure for estimating the covariance matrix of the coefficients should be applied.

5. ASYMPTOTIC PROPERTIES

In matrix format, the original model can be rewritten as

$$ y = X\beta + Z\gamma + \eta \rho + \varepsilon $$

(5.1)

with $y = (y_1, \ldots, y_n)'$, $X = (x_1, \ldots, x_n)'$, $Z = (z_1, \ldots, z_n)'$, $\eta = (\eta_1, \ldots, \eta_n)'$, and $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)'$. The vector $\eta$ is unknown and is replaced by the estimator $\hat{\eta} = (\hat{\eta}_1, \ldots, \hat{\eta}_n)'$ as proposed in Section 3. This leads to

$$ y = W\alpha + \hat{\eta} \rho + (\eta - \hat{\eta}) \rho + \varepsilon $$

(5.2)

with $W = (X,Z)$ and $\alpha = (\beta', \gamma)'$. Accordingly, the endogeneity corrected estimator is obtained as

$$ \hat{\alpha} = (W'M_{\hat{\eta}}W)^{-1}W'M_{\hat{\eta}}y $$

with the residual maker matrix $M_{\hat{\eta}}$ as defined in the previous section.
We are interested in the asymptotic properties of

$$\sqrt{n} (\hat{\alpha} - \alpha) = \sqrt{n}(W'M_\eta W)^{-1} W'M_\eta \tilde{\varepsilon}$$

$$= \left(\frac{1}{n} W'M_\eta W\right)^{-1} \frac{1}{\sqrt{n}} W'M_\eta \varepsilon + \left(\frac{1}{n} W'M_\eta W\right)^{-1} \frac{1}{\sqrt{n}} W'M_\eta (\eta - \hat{\eta})\rho . \quad (5.3)$$

The proofs for all results are deferred to the appendix.

We start by analyzing the difference between $\hat{\eta}_i$ and $\eta_i$, where the following assumption is employed:

**Assumption 1.**

(i) The $k \times 1$ vector $x_i$ is i.i.d. with $E(x_i) = \mu_x$ and $E(x_i x_i') = \Sigma_x$, where $\Sigma_x$ is a positive definite matrix.

(ii) There exists a decomposition of the endogenous regressor such that $z_i = \delta' x_i + e_i$, where $e_i$ is i.i.d. with a strictly monotonously increasing c.d.f.

(iii) There exists a strictly monotonous function $f(\cdot)$ such that the vector $(f(e_i), \varepsilon_i)$ is distributed as

$$\begin{pmatrix} f(e_i) \\ \varepsilon_i \end{pmatrix} \overset{iid}{\sim} \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_f & 0 \\ 0 & \sigma^2_\varepsilon \end{pmatrix}\right).$$

(iv) $\varepsilon_i$ is independent of $x_i$ and $e_i$.

(v) $\frac{1}{n} W'W \xrightarrow{p} Q$, where $Q$ is a finite, positive definite matrix with $k + 1$ rows and columns.

(vi) The vector $\eta$ does not lie in the column space of $W$.

The following lemma is the basis for the asymptotic analysis.

**Lemma 1.** Under Assumption 1, it holds that $\hat{\eta}_i - \eta_i = O_p(n^{-1/2})$ for every $i$ as $n \to \infty$.

Let $M_\eta = I_n - \eta \eta' / \eta' \eta.$ Using the previous lemma we are able to show
Lemma 2. Under Assumption 1, \( \frac{1}{n} W' M_\eta W - \frac{1}{n} W' M_\eta W \xrightarrow{p} 0 \) and \( \frac{1}{n} W' M_\eta W \xrightarrow{p} S \), where \( S \) is a finite, positive definite matrix.

Let us now consider the term
\[
\frac{1}{\sqrt{n}} W' M_\hat{\eta} \varepsilon = \frac{1}{\sqrt{n}} W' \varepsilon - \frac{1}{\sqrt{n}} W' \hat{\eta}' \varepsilon \frac{1}{\hat{\eta}' \hat{\eta}}.
\]
We have

Lemma 3. Under Assumptions 1, \( \frac{1}{\sqrt{n}} W' M_\hat{\eta} \varepsilon \xrightarrow{d} N(0, \sigma_\varepsilon^2 S) \).

Collecting these results we obtain for the first sum in (5.3)
\[
\left( \frac{1}{n} W' M_\eta W \right)^{-1} \frac{1}{\sqrt{n}} W' M_\hat{\eta} \varepsilon = \left( \frac{1}{n} W' M_\eta W \right)^{-1} \frac{1}{\sqrt{n}} W' M_\eta \varepsilon + o_p(1)
\]
\[
\xrightarrow{d} N(0, \sigma_\varepsilon^2 S^{-1}).
\]

Note that the same limiting distribution results if the regression is augmented with \( \eta_i \) instead of \( \hat{\eta}_i \) and, therefore, the second sum in (5.3) captures the effect on the estimation error \( \hat{\eta} - \eta \).

Let us now analyze the term
\[
\frac{1}{\sqrt{n}} W' M_\hat{\eta} (\eta - \hat{\eta}) = \frac{1}{\sqrt{n}} W'(\eta - \hat{\eta}) - \left( \frac{1}{n} W' \hat{\eta} \right) \left( \frac{1}{\sqrt{n}} \hat{\eta}' (\eta - \hat{\eta}) \right) / (n^{-1} \hat{\eta}' \hat{\eta}).
\]
The main tool for analyzing the second summand is Lemma 4. Its arguably most interesting aspect is that the limit distribution does not depend on the model.

Lemma 4. Under Assumption 1, \( G_n := \frac{1}{\sqrt{n}} \hat{\eta}' (\eta - \hat{\eta}) \xrightarrow{d} G \xrightarrow{} \mathcal{N}(0, 1/2) \).

The results obtained so far do not depend on whether we use the residuals of the regression of \( z \) on \( x \) or use \( z \) directly. In the latter case (that is by assuming \( \delta = 0 \) in (2.7)), it is possible to derive the asymptotic normality of \( \sqrt{n} (\hat{\alpha} - \alpha) \). In particular, it is then possible to derive the asymptotic distribution of the term \( \frac{1}{\sqrt{n}} W'(\eta - \hat{\eta}) \), which appears to be much more cumbersome if \( x \) and \( \eta - \hat{\eta} \) are not independent. This term is the only term in the
limiting expression whose distribution we cannot approximate by simple plug-in estimators as it depends on the unknown function \( f \) from (2.2). Therefore we first focus on this case and introduce

**Assumption 2.** Assume that

1. \( \delta = 0 \) in (2.7) and, therefore, \( \hat{\eta}_i \) is constructed by letting \( \hat{e}_i = z_i \).

2. The integrals in (A.2) exist and are finite.

Note that the second part of this assumption potentially restricts the class of distributions that \( e_i \) can have. Numerical evidence suggests that the assumption is fulfilled for the distributions considered in the simulation study below and e.g. for the \( t_4 \)-distribution, but not for the \( t_3 \)-distribution, for example. It would be possible to proceed without this assumption if the \( \hat{\eta}_i \) are truncated from below and above, but we do not follow this path, as we do not want to introduce a tuning parameter.

**Lemma 5.** Under Assumptions 1 and 2 we have \( H_n := \frac{1}{\sqrt{n}} W'(\eta - \hat{\eta}) \overset{d}{\rightarrow} \mathcal{N}(0, V_H) \), where the matrix \( V_H \) might not be positive definite.

Summing up the previous results, we obtain

**Theorem 1.** Let Assumptions 1 be fulfilled. Then it holds that

1. \( \hat{\alpha} \) is a consistent estimator for \( \alpha \).

2. If in addition Assumption 2 is fulfilled, \( \sqrt{n}(\hat{\alpha} - \alpha) \overset{d}{\rightarrow} \mathcal{N}(0, V_\alpha) \) for some positive definite covariance matrix \( V_\alpha \).

The asymptotic variance of \( \hat{\alpha} \) depends on the data generating process and is difficult to estimate as the term from Lemma 5 depends on the unknown function \( f(\cdot) \). This is the reason why we approximate the distribution of \( \sqrt{n}(\hat{\alpha} - \alpha) \) by using bootstrap standard errors.
Here, we simply generate bootstrap samples by drawing with replacement from the original data, as described in the end of Section 3.

An even simpler approach is to ignore the estimation error of $\hat{\eta}_i$ and considering the usual OLS standard errors from the augmented regression. Although this does not give the asymptotically correct size (of the $t$ or $F$-tests and the confidence intervals for $\alpha$), but simulations indicate that the error can be assumed to be small.

Finally we analyze the asymptotic distribution of the exogeneity test for the null hypothesis $\rho = 0$ in the regression (5.1). Our previous results might suggest that the estimation error $\hat{\eta} - \eta$ affects the distribution of the $t$ statistic. As we show in the next theorem, however, this is not the case and under the null hypothesis standard statistical inference applies.

**Theorem 2.** Let Assumptions 4 be fulfilled. Under $H_0: \rho = 0$ the regressor $z_i$ is exogenous and the respective $t$-statistic has a standard normal limiting distribution as $n \to \infty$.

6. SMALL SAMPLE PROPERTIES

In order to investigate the small sample properties of our estimator, we consider a simple data generating process given by

$$ y_i = \beta_0 + \beta_1 x_i + \gamma z_i + u_i $$

where $(\beta_0, \beta_1, \gamma)' = (0, 1, 1)'$, $x_i \overset{iid}{\sim} \mathcal{N}(0, 1)$ and $z_i = \delta x_i + e_i$. We consider two different distributions of the component $e_i$. Generating $e_i$ as a $\Gamma(1, 1)$ distributed random variable, the distribution of $e_i$ (and therefore $z_i$) becomes highly skewed, see Figure 6.1. Accordingly, this setup is favorable for identifying the endogeneity in the error term as the distributions of the error and the endogenous regressor are quite different. A more challenging situation results if $e_i$ is generated as a $\Gamma(3, 2)$ distribution as this distribution is more similar to the normal distribution (see Figure 6.1 for the density of this distribution). For all Monte Carlo
In our experiments we set $n = 200$.

The error term is generated as $u_i = \rho \eta_i + \varepsilon_i$, where $\varepsilon_i \overset{iid}{\sim} \mathcal{N}(0, 1)$ and the standard normal distributed random variable $\eta_i$ is generated as $\eta_i = \Phi^{-1}[F_\Gamma(e_i)]$ and $F_\Gamma(e_i)$ is the c.d.f. of the respective $\Gamma$-distribution. We distinguish the cases of uncorrelated ($\delta = 0$) and correlated ($\delta = 1$) regressors as well as of no ($\rho = 0$), moderate ($\rho = 0.5$) and strong ($\rho = 0.9$) endogeneity. The proposed nonparametric control function estimator ("np-CF") is compared with OLS (i.e. ignoring the endogeneity of $z_i$), with the estimator by Park and Gupta (2012) ("G-P") and with the "2sCOPE" estimator by Yang et al. (2022). In order to investigate the validity of statistical inference we consider $t$-type test statistics for the hypotheses $\beta = 1$ and $\gamma = 1$ respectively. For the OLS estimator the usual $t$-statistics are used, whereas the other tests are performed using bootstrap inference. Our experience suggests that using 100 bootstrap replications is typically enough to obtain reliable estimates for the unknown standard deviations of the coefficients.

---

4 We employ the R-package ‘REndo’ for computing this estimator.
Table 1: Simulation results for $z_i \sim \Gamma(1,1)$

### Uncorrelated regressors: $\delta = 0$

| $\rho = 0$ | OLS | G-P | np-CF | 2sCOPE | OLS | G-P | np-CF | 2sCOPE |
|------------|-----|-----|-------|--------|-----|-----|-------|--------|
| bias       | -0.002 | -0.002 | -0.001 | -0.001 | 0.003 | -0.001 | -0.001 | -0.001 |
| st.err.    | 0.070  | 0.070 | 0.074 | 0.071 | 0.073  | 0.194 | 0.180 | 0.177  |
| RMSE       | 0.070  | 0.070 | 0.074 | 0.071 | 0.073  | 0.194 | 0.180 | 0.177  |
| rej. $H_0$ | 0.046  | 0.052 | 0.040 | 0.049 | 0.058  | 0.044 | 0.048 | 0.059  |

| $\rho = 0.5$ | OLS | G-P | np-CF | 2sCOPE | OLS | G-P | np-CF | 2sCOPE |
|--------------|-----|-----|-------|--------|-----|-----|-------|--------|
| bias         | 0.000  | 0.000 | -0.000 | 0.000  | 0.454  | -0.009 | 0.016 | 0.007  |
| st.err.      | 0.061  | 0.064 | 0.089 | 0.072 | 0.064  | 0.168 | 0.154 | 0.151  |
| RMSE         | 0.061  | 0.064 | 0.089 | 0.072 | 0.458  | 0.168 | 0.155 | 0.151  |
| rej. $H_0$   | 0.027  | 0.059 | 0.041 | 0.051 | 1.000  | 0.033 | 0.042 | 0.049  |

| $\rho = 0.9$ | OLS | G-P | np-CF | 2sCOPE | OLS | G-P | np-CF | 2sCOPE |
|--------------|-----|-----|-------|--------|-----|-----|-------|--------|
| bias         | -0.001 | -0.001 | 0.004 | 0.002  | 0.814  | -0.024 | 0.013 | -0.001 |
| st.err.      | 0.032  | 0.041 | 0.114 | 0.070 | 0.048  | 0.126 | 0.116 | 0.116  |
| RMSE         | 0.032  | 0.041 | 0.114 | 0.070 | 0.815  | 0.128 | 0.117 | 0.116  |
| rej. $H_0$   | 0.000  | 0.037 | 0.068 | 0.045 | 1.000  | 0.016 | 0.066 | 0.060  |

### Correlated regressors: $\delta = 1$

| $\rho = 0.5$ | OLS | G-P | np-CF | 2sCOPE | OLS | G-P | np-CF | 2sCOPE |
|--------------|-----|-----|-------|--------|-----|-----|-------|--------|
| bias         | -0.474 | -0.461 | -0.015 | -0.161 | 0.648  | 0.501 | 0.023 | 0.220  |
| st.err.      | 0.097  | 0.095 | 0.191 | 0.304 | 0.096  | 0.440 | 0.231 | 0.417  |
| RMSE         | 0.484  | 0.471 | 0.192 | 0.344 | 0.655  | 0.667 | 0.233 | 0.472  |
| rej. $H_0$   | 0.995  | 0.996 | 0.044 | 0.064 | 1.000  | 0.358 | 0.051 | 0.069  |

| $\rho = 0.9$ | OLS | G-P | np-CF | 2sCOPE | OLS | G-P | np-CF | 2sCOPE |
|--------------|-----|-----|-------|--------|-----|-----|-------|--------|
| bias         | -0.846 | -0.822 | -0.013 | -0.259 | 1.158  | 0.853 | 0.024 | 0.357  |
| st.err.      | 0.076  | 0.082 | 0.160 | 0.254 | 0.081  | 0.437 | 0.164 | 0.343  |
| RMSE         | 0.850  | 0.826 | 0.161 | 0.363 | 1.161  | 0.959 | 0.166 | 0.496  |
| rej. $H_0$   | 1.000  | 1.000 | 0.063 | 0.215 | 1.000  | 0.711 | 0.070 | 0.220  |

**Note:** bias, standard errors (st.err.) and root mean squared error (RMSE) based on 1000 replications. “rej. $H_0$”: empirical sizes of the (bootstrap) tests for $H_0 : \beta = 1$ resp. $\gamma = 1$. The sample size is $n = 200$. 
Table 2: Simulation results for $z_i \sim \Gamma(3, 2)$

### Uncorrelated regressors: $\delta = 0$

| $\rho$ = 0 | OLS | G-P | $\beta$ | np-CF | 2sCOPE | OLS | G-P | $\gamma$ | np-CF | 2sCOPE |
|------------|-----|-----|---------|-------|--------|-----|-----|---------|-------|--------|
| bias       | 0.000 | 0.000 | 0.000   | 0.000 | -0.000 | 0.014 | 0.021 | 0.020 |
| st.err.    | 0.073 | 0.073 | 0.077   | 0.076 | 0.072  | 0.295 | 0.281 | 0.280 |
| RMSE       | 0.073 | 0.073 | 0.077   | 0.076 | 0.072  | 0.296 | 0.282 | 0.280 |
| rej. $H_0$ | 0.061 | 0.065 | 0.035   | 0.046 | 0.055  | 0.042 | 0.037 | 0.040 |

| $\rho$ = 0.5 | OLS | G-P | $\beta$ | np-CF | 2sCOPE | OLS | G-P | $\gamma$ | np-CF | 2sCOPE |
|--------------|-----|-----|---------|-------|--------|-----|-----|---------|-------|--------|
| bias         | 0.002 | 0.002 | 0.004   | 0.003 | 0.479  | 0.088 | 0.000 | 0.003 |
| st.err.      | 0.064 | 0.065 | 0.079   | 0.074 | 0.065  | 0.275 | 0.262 | 0.259 |
| RMSE         | 0.064 | 0.065 | 0.079   | 0.074 | 0.483  | 0.289 | 0.262 | 0.259 |
| rej. $H_0$   | 0.030 | 0.072 | 0.046   | 0.052 | 1.000  | 0.078 | 0.036 | 0.043 |

| $\rho$ = 0.9 | OLS | G-P | $\beta$ | np-CF | 2sCOPE | OLS | G-P | $\gamma$ | np-CF | 2sCOPE |
|--------------|-----|-----|---------|-------|--------|-----|-----|---------|-------|--------|
| bias         | -0.000 | -0.000 | 0.002   | 0.001 | 0.869  | 0.172 | 0.021 | 0.021 |
| st.err.      | 0.032 | 0.035 | 0.082   | 0.072 | 0.051  | 0.197 | 0.181 | 0.183 |
| RMSE         | 0.032 | 0.035 | 0.082   | 0.072 | 0.871  | 0.262 | 0.182 | 0.184 |
| rej. $H_0$   | 0.000 | 0.036 | 0.065   | 0.063 | 1.000  | 0.201 | 0.085 | 0.096 |

### Correlated regressors: $\delta = 1$

| $\rho$ = 0.9 | OLS | G-P | $\beta$ | np-CF | 2sCOPE | OLS | G-P | $\gamma$ | np-CF | 2sCOPE |
|--------------|-----|-----|---------|-------|--------|-----|-----|---------|-------|--------|
| bias         | -0.563 | -0.561 | -0.019  | -0.383 | 0.742  | 0.767 | 0.027 | 0.504 |
| st.err.      | 0.097 | 0.097 | 0.319   | 0.452 | 0.102  | 0.411 | 0.408 | 0.593 |
| RMSE         | 0.571 | 0.569 | 0.319   | 0.593 | 0.749  | 0.870 | 0.409 | 0.778 |
| rej. $H_0$   | 1.000 | 1.000 | 0.055   | 0.178 | 1.000  | 0.599 | 0.051 | 0.175 |

| $\rho$ = 0.9 | OLS | G-P | $\beta$ | np-CF | 2sCOPE | OLS | G-P | $\gamma$ | np-CF | 2sCOPE |
|--------------|-----|-----|---------|-------|--------|-----|-----|---------|-------|--------|
| bias         | -1.010 | -1.007 | -0.032  | -0.693 | 1.328  | 1.348 | 0.044 | 0.910 |
| st.err.      | 0.067 | 0.067 | 0.226   | 0.322 | 0.088  | 0.276 | 0.277 | 0.413 |
| RMSE         | 1.012 | 1.009 | 0.228   | 0.764 | 1.331  | 1.376 | 0.281 | 1.000 |
| rej. $H_0$   | 1.000 | 1.000 | 0.091   | 0.683 | 1.000  | 0.981 | 0.081 | 0.674 |

**Note:** see Table 1.
We first present the results for $\delta = 0$ in Table I, that is, the setup with uncorrelated regressors that is supposed by Park and Gupta (2012). Not surprisingly, OLS performs best in case of no endogeneity with a RMSE less than the half of the other three approaches. This highlights the efficiency loss resulting from the endogeneity correction if it is in fact not necessary. A similar efficiency loss occurs when employing external instruments.

Note that in this case the asymptotic distribution of the latter IV estimator with $v_i$ as external instrument is given by

$$\sqrt{n}(\hat{\gamma}_{iv} - \gamma) \xrightarrow{d} \mathcal{N}(0, V_{ols}/r_{zv}^2).$$

where $V_{ols}$ is the asymptotic variance of the OLS estimator and $r_{zv}^2$ is the $R^2$ from the (first stage) regression of $z_i$ on $v_i$. Accordingly, a comparable loss of efficiency results from an IV estimator if the first stage $R^2$ corresponds to $(0.073/0.18)^2 = 0.165$.

In the presence of endogeneity (i.e. $\rho \neq 0$) the OLS estimator of the parameter $\gamma$ is highly biased. Note that for the case of uncorrelated regressors, the coefficient $\beta$ can be estimated unbiasedly even if the regressor $z_i$ is endogenous. Accordingly, all estimators (including OLS) for $\beta$ are unbiased in this case. On the other hand, the OLS estimator for $\gamma$ is severely biased as the corresponding regressor is endogenous. The copula based endogeneity corrections $G$-$P$ and $2sCOPE$ and our $np$-$CF$ estimator effectively remove the bias and perform more or less similarly. Furthermore, the empirical sizes of the (bootstrap) $t$-tests are close to 0.05.

The case of correlated regressors with $\delta = 1$ is considered in the second panel of Table 1. As expected, in this case the OLS estimator is biased for both coefficients $\beta$ and $\gamma$. The bias of the $G$-$P$ estimator is similarly to the OLS bias for $\beta$ and only slightly smaller for $\gamma$. A substantial bias reduction is obtained by applying the $2sCOPE$ estimator but some bias remains in particular if $\rho = 0.9$. This is due to the fact that the $2sCOPE$ estimator assumes

---

5 It should be noted, however, that it is easy to verify that the endogeneity correction is unnecessary by observing an insignificant $t$-statistic for the correction term $\tilde{\eta}$. 

17
a particular nonlinear dependence between $x_i$ and $z_i$, whereas in our data generating process the dependence between $z_i$ and $x_i$ is linear. The np-CF estimator is the only estimator that efficiently removes the bias from both parameters.

The results for the second Monte Carlo experiment is presented in Table 2. In this case the endogeneity component $e_i$ is $\Gamma(3, 2)$ distributed. As this distribution comes fairly close to a normal distribution, we expect that the endogeneity corrections have problems to empirically identify the endogenous component of the error term. Indeed, the standard errors of the endogeneity corrected standard errors for the estimators of $\gamma$ increase about 50 percent. Apart from the higher standard errors the findings are qualitatively similar to the results presented in Table 1. It is interesting to note, however, that the magnitude of the bias of the 2sCOPE estimator comes close to the bias of the OLS estimator suggesting that this estimator is not able to cope with the (linear) correlation between the regressors $x_i$ and $z_i$.

7. EMPIRICAL APPLICATION

In this section, we apply our new estimator to empirical wage data, thereby revisiting the classical economic problem of estimating the returns to schooling (see e.g. Harmon, Oosterbeek, and Walker, 2000). Similarly as in Chernozhukov, Fernández-Val, and Melly (2013) and Rothe and Wied (2013), we analyze micro-level data from the US Current Population Survey in 1988. The sample size is $n = 144,750$ and the model is given by

\[
\log(wage_i) = \beta_0 + \gamma \text{educ}_i + \beta_1 \text{exper}_i + \beta_2 \text{exper}^2_i + \beta_3 \text{married}_i + \beta_4 \text{parttime}_i + \beta_5 \text{union}_i + \beta_6 \text{smsa}_i + \beta_7 \text{nonwhite}_i + u_i
\]

Accordingly we consider a linear regression in which the logarithm of hourly wages for individual $i$ is explained by the years of education and different control variables. The latter consist of the years of working experience (linearly and quadratically) and several binary variables (married, working in part time, member in a union, living in a standard metropolitan
statistical area, being non-white). The control variables are assumed to be exogenous. The years of education is assumed to be an endogenous variable as it might be correlated with unobservable worker’s characteristics in the error term $u_i$ such as ability or motivation. The key goal is to find a reliable estimate for $\gamma$.

The literature in labor economics discusses several approaches for estimating $\gamma$ and provides different empirical results for different datasets. For example, if $educ_i$ is instrumented with family characteristics such as the years of education of the parents, the IV estimates are often smaller than the OLS estimates (see Wooldridge 2009). This fits to the economic intuition that years of education might be positively correlated with ability/motivation. If institutional characteristics are used as instruments, the IV estimates are often larger (see Card 2000). Lemke and Rischall (2003) try to merge these results with the conclusion that the OLS estimators might be biased upwards.

This analysis contributes to this discussion by providing results of a consistent estimator for $\gamma$ that do not depend on external instruments. The starting point is the observation that the distribution of the regressor $educ$ is obviously not normal, as the histogram in Figure 7.2 (left) shows. On the one hand, the distribution is discrete and on the other hand, it is highly skewed. So, we can suitably augment the regression (7.1) by the normalizing transformation of $educ$ as discussed earlier in the paper. In the first step, residuals $\hat{e}_i$ from an OLS regression of $educ$ on all exogenous control variables are obtained. In the second step, the augmented regressor $\hat{n}_i = \Phi^{-1} \frac{n}{n+0.5} \hat{F}_i(\hat{e}_i)$ is calculated. The kernel density estimate of the distribution of the $\hat{n}_i$ in Figure 7.2 (right) indicates that the distribution is normal. In particular, the distribution appears to be much more continuous than that of $educ$ as there are different realizations for the different combinations of $educ$ and the control variables. It is the difference between the two distributions which enables identification of $\gamma$.

Table 3 presents the results (estimates, standard errors and $t$-statistics) of the standard OLS
Figure 7.2: Comparison of the distributions of $educ$ and its normalized transformation

estimation of model (7.1) and the results of the robustified estimation, where $\hat{\eta}_i$ is included. In the latter case, we report both the bootstrap standard errors and $t$-statistics as well as the OLS $t$-statistics. In the latter case, the estimation error for $\hat{\eta}_i$ is ignored, but the differences are small and not systematic. Due to the large sample size, the $p$-values for all coefficients are smaller than $2 \cdot 10^{-16}$. One exception is the coefficient $\rho_{\eta}$ for $\hat{\eta}_i$ which is somewhat larger with $2.5 \cdot 10^{-5}$. This reflects that, on the one hand, all control variables were used to calculate $\hat{\eta}_i$, but, on the other hand, $\hat{\eta}_i$ still provides considerable explanatory power. The latter observation implies that it is reasonable to assume that the regressor $educ$ is endogenous.

In the case of OLS, we have $\hat{\gamma} = 0.083$, whereas we have $\hat{\gamma} = 0.073$ when correcting for endogeneity. So, we observe a slight reduction of the estimated return for education which is what [Lemke and Rischall (2003)] also suggest. The other coefficients remain very similar, which could have been expected as the residuals $\hat{e}_i$ are orthogonal to the exogenous regressors.

For all coefficients, the $t$-statistics in absolute values are lower and the standard errors are higher in the augmented regression compared to OLS. This is in particular the case for the intercept and $\gamma$ and can be explained with the high correlation (0.91) of $educ$ and $\hat{\eta}_i$. A similar phenomenon is observed in other studies of endogeneity-corrected estimators such as [Park and Gupta (2012)].
Table 3: Estimation results for the model (7.1), standard errors and $t$-statistics

| Coefficient | OLS | Augmented |
|-------------|-----|-----------|
|              | Estimate | st.err. | $t$-stat. | Estimate | st.err. BS | $t$-stat. BS | $t$-stat. OLS |
| $\gamma$    | $0.083$ | $< 0.001$ | $81.44$ | $0.072$ | $0.003$ | $25.01$ | $28.98$ |
| $\beta_0$   | $0.559$ | $0.007$ | $180.31$ | $0.699$ | $0.037$ | $18.59$ | $21.85$ |
| $\beta_1$   | $0.027$ | $< 0.001$ | $78.71$ | $0.028$ | $< 0.001$ | $85.52$ | $74.31$ |
| $\beta_2$   | $0.001$ | $< 0.001$ | $-54.14$ | $< 0.001$ | $< 0.001$ | $-49.39$ | $-43.15$ |
| $\beta_3$   | $0.233$ | $0.003$ | $74.91$ | $0.234$ | $0.003$ | $86.50$ | $74.85$ |
| $\beta_4$   | $0.119$ | $0.003$ | $46.09$ | $0.124$ | $0.003$ | $38.48$ | $42.62$ |
| $\beta_5$   | $-0.333$ | $0.003$ | $-107.23$ | $-0.341$ | $0.003$ | $-98.85$ | $-95.53$ |
| $\beta_6$   | $-0.114$ | $0.003$ | $-32.79$ | $-0.118$ | $0.004$ | $-33.75$ | $-32.51$ |
| $\beta_7$   | $0.181$ | $0.003$ | $69.72$ | $0.188$ | $0.003$ | $64.95$ | $61.19$ |
| $\rho_\eta$ | $-$     | $-$     | $-$     | $0.026$ | $0.007$ | $3.85$  | $4.21$  |
| $R^2$       | $0.415$ | $-$     | $-$     | $0.415$ | $-$     | $-$     | $-$     |

Note: BS: computed based on 100 bootstrap replications. OLS: regression $t$-statistic ignoring the estimation error in $\eta_i$.

8. CONCLUDING REMARKS

This paper proposes a method for solving the endogeneity problem without external instruments or specifying a copula. The endogeneity correction is obtained by simply augmenting the regression with a conformably transformed regressor. We study the asymptotic properties and analyse the small sample properties of the resulting estimator. Our results suggest that the proposed estimation framework may provide a useful tool for estimating regression models with endogenous regressors if no suitable instruments are available. Even if instrumental variables are at hand, additional nonlinear instruments can be constructed for improve the efficiency of the IV estimator. Furthermore, the nonparametric control function may be employed for testing the exogeneity of the regressors in empirical situations where valid instruments are missing.

An important drawback of our approach is however that the distribution of the endogenous regressor needs to be substantially different from normality. If the distribution comes close to the normal distribution, then the estimator suffers from multicollinearity among the
endogenous regressor and the correction term resulting in potentially large standard errors. It therefore remains to see whether the alternative endogeneity correction turns out to useful in empirical practice.

REFERENCES

ANGRIST, J.-D. AND J.-S. PISCHKE (2008): Mostly Harmless Econometrics: An Empiricist’s Companion, Princeton University Press.

CARD, D. (2000): “The Causal Effect of Earnings on Education,” in Handbook of Labor Economics, Volume 3, ed. by O. Ashenfelter and D. Card, Elsevier, Amsterdam, 1802–1863.

CHERNOZHUKOV, V., I. FERNÁNDEZ-VAL, AND B. MELLY (2013): “Inference on counterfactual distributions,” Econometrica, 81, 2205–2268.

HARMON, C., H. OOSTERBEEK, AND I. WALKER (2000): “The Returns to Education: A Review of Evidence, Issues and Deficiencies in the Literature,” Centre for the Economics of Education, LSE, CEE Discussion Papers.

HASCHKA, R. (2021): “Exploiting Between-Regressor Correlation to Robustify Copula Correction Models for Handling Endogeneity,” mimeo.

——— (2022): “Handling Endogenous Regressors Using Copulas: A Generalization to Linear Panel Models With Fixed Effects and Correlated Regressors,” Journal of Marketing Research, forthcoming.

KOENKER, R. AND Z. XIAO (2002): “Inference on the Quantile Regression Process,” Econometrica, 70, 1583–1612.

LEMKE, R. AND I. RISCHALL (2003): “Skill, Parental Income, and IV Estimation of the Returns to Schooling,” Applied Economics Letters, 10, 281–236.
PARK, S. AND S. GUPTA (2012): “Handling Endogenous Regressors by Joint Estimation Using Copulas,” *Marketing Science*, 31, 567–586.

ROTHE, C. AND D. WIED (2013): “Misspecification Testing in a Class of Conditional Distributional Models,” *Journal of the American Statistical Association*, 108, 314–324.

STOCK, J. AND M. YOGO (2005): “Testing for Weak Instruments in Linear IV Regression,” in *Identification and Inference for Econometric Models*, ed. by D. Andrews, Cambridge University Press, New York, 80–108.

WEBEL, K. AND D. WIED (2016): *Stochastische Prozesse*, Springer-Gabler.

WOOLDRIDGE, J. (2009): *Introductory Econometrics*, South-Western.

——— (2015): “Control Function Methods in Applied Econometrics,” *Journal of Human Resources*, 50, 420–445.

YANG, F., Y. QIAN, AND H. XIE (2022): “Addressing Endogeneity Using a Two-stage Copula Generated Regressor Approach,” *NBER Working Paper Series*. 
Proof of Lemma I

An application of the delta method yields

\[ \hat{\eta}_i - \eta_i = \Phi^{-1} \left[ \frac{n}{n+0.5} \hat{F}_e(\hat{e}_i) \right] - \Phi^{-1} [F_e(e_i)] \]

\[ = \frac{1}{\phi(\Phi^{-1}(F_e(e_i)))} \left[ \frac{n}{n+0.5} \hat{F}_e(\hat{e}_i) - F_e(e_i) \right] + o_p \left( \frac{n}{n+0.5} \hat{F}_e(\hat{e}_i) - F_e(e_i) \right). \]

Consider \( \hat{F}_e(\hat{e}_i) \). It holds that \( \hat{e} = (\hat{e}_1, \ldots, \hat{e}_n)' = M_X Z = M_X e \) for \( e = (e_1, \ldots, e_n)' \). So, \( \hat{e}_i = e_i - \eta_i' X(X'X)^{-1}X'e \), where \( \eta_i \) is the \( i \)-th unit vector. Moreover,

\[ \hat{F}_e(\hat{e}_i) = \frac{1}{n} \sum_{j=1}^{n} I(\hat{e}_j \leq \hat{e}_i) = \frac{1}{n} \sum_{j=1}^{n} I(e_j - \eta_j' X(X'X)^{-1}X'e \leq e_i - \eta_i' X(X'X)^{-1}X'e) \]

\[ = \frac{1}{n} \sum_{j=1}^{n} I(e_j \leq e_i + (\eta_i - \eta_j)' X(X'X)^{-1}X'e) \]

Next, we rewrite the difference \( \hat{F}_e(\hat{e}_i) - F_e(e_i) \) as

\[ \hat{F}_e(\hat{e}_i) - F_e(e_i) = \frac{1}{n} \sum_{j=1}^{n} I(e_j \leq e_i + (\eta_j - \eta_i)' X(X'X)^{-1}X'e) - F_e(e_i) \]

\[ = \frac{1}{n} \sum_{j=1}^{n} \left[ I(e_j \leq e_i + (\eta_j - \eta_i)' X(X'X)^{-1}X'e) - F_e(e_i + (\eta_i - \eta_j)' X(X'X)^{-1}X'e) \right] \]

\[ + \frac{1}{n} \sum_{j=1}^{n} \left[ F_e(e_i + (\eta_i - \eta_j)' X(X'X)^{-1}X'e) - F_e(e_i) \right] \]

\[ := f_{i,n} + g_{i,n}. \]

Consider the first and second moments of \( \sqrt{n}f_{i,n} \). Denote

\[ f_{i,j,n} := I(e_j \leq e_i + (\eta_j - \eta_i)' X(X'X)^{-1}X'e) - F_e(e_i + (\eta_j - \eta_i)' X(X'X)^{-1}X'e). \]

With Assumption I (stochastic boundedness of \( X \)), it holds that

\[ (\eta_j - \eta_i)' X(X'X)^{-1}X'e = O_p(1/\sqrt{n}). \] (A.1)

By applying the law of iterated expectations on each summand (conditioning on \( X \) and all components of \( e \) except of \( e_j \)) we then obtain
\(E(f_{i,n}) = O_p(1/\sqrt{n})\). Moreover, note that

\[Var(\sqrt{n}f_{i,n}) = \frac{1}{n} \sum_{j=1}^{n} \sum_{k=1}^{n} Cov(f_{i,j,n}, f_{i,k,n}).\]

With (A.1), it holds \(f_{i,j,n} = O(1/\sqrt{n})\). Together with the fact that the \(f_{i,j,n}\) are identically distributed for fixed \(i\), we obtain \(Var(\sqrt{n}f_{i,n}) = O(1)\) such that \(f_{i,n} = O_p(n^{-1/2})\).

Consider now the term \(g_{i,n}\). By another application of the delta method, \(g_{i,n}\) can be expressed as

\[
\frac{1}{n} \sum_{j=1}^{n} f(e_i)(t_j - u_i)'X(X'X)^{-1}X'e + O_p(n^{-3/2}),
\]

which is also \(O_p(n^{-1/2})\). Observing that \(\frac{n}{n+0.5} \to 1\) the proof is finished.

**Proof of Lemma 2**

Under Assumption 1.2, it holds \(\frac{1}{n} W'W \overset{P}{\to} Q\). Define the \(k+1\) dimensional vector \(a_n := \frac{1}{n} W'\hat{\eta}\).

The first \(k\) elements of the vector are comprised in the term \(a_{1n} := \frac{1}{n} \sum_{i=1}^{n} x_i \eta_i\) and the last element is \(a_{2n} := \frac{1}{n} Z'\hat{\eta}\). It holds \(a_{1n} = \frac{1}{n} \sum_{i=1}^{n} x_i \eta_i\). For \(a_{1n}\), consider the decomposition

\[a_{1n} = \frac{1}{n} \sum_{i=1}^{n} x_i \eta_i + \frac{1}{n} \sum_{i=1}^{n} x_i (\hat{\eta}_i - \eta_i) =: a_{1n,1} + a_{1n,2}.
\]

Then, \(a_{1n,1}\) converges to zero in probability due to the independence of \(x_i\) and \(\eta_i\) and \(E(\eta_i) = 0\) (which yields \(E(x_i \eta_i = 0)\)). With the Cauchy-Schwarz inequality, we have for the \(j^\text{th}\) component of \(a_{1n,2}\)

\[
||a_{1n,2}|| = \left|\left| \frac{1}{n} \sum_{i=1}^{n} x_i (\hat{\eta}_i - \eta_i) \right|\right| \leq \sqrt{\frac{1}{n} \sum_{i=1}^{n} ||x_i||^2} \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\hat{\eta}_i - \eta_i)^2}.
\]

With Assumption 2.1, it holds that \(\frac{1}{n} \sum_{i=1}^{n} ||x_i||^2 = O_p(1)\).

Now consider the term

\[a_{1n,3} := \frac{1}{n} \sum_{i=1}^{n} (\hat{\eta}_i - \eta_i)^2.
\]

It holds that \(a_{1n,3} = \frac{1}{n} \sum_{i=1}^{n} \left( \Phi^{-1}(F_e(e_i)) - \Phi^{-1}[\tilde{F}_e(\tilde{e}_i)] \right)^2 + o_p(1)\). Note that \(F_e(e_i)\) is stan-
standard uniformly distributed and \( \hat{F}_c(e) \) is discretely distributed over the set \{1/n, \ldots, n/n\}. Furthermore, under Assumption 1 \( \hat{\rho}_i = \frac{n}{n+0.5} \hat{F}_c(e_i) \) is consistent for \( \eta_i \) for all \( i \) with Lemma 1.

It follows that there is an ordering of the indices such that \( \hat{F}_c(e_i) = i/n \) and \( F_c(e_i) = F_n^{-1}(i/n) \) with the empirical distribution function of a standard uniform distribution \( F_n \). This means that we have with \( \tilde{n} := n + 0.5 \)

\[
a_{1n,3} = \frac{1}{n} \sum_{i=1}^{n} \left( \Phi^{-1}(F_n^{-1}(i/\tilde{n})) - \Phi^{-1}(i/\tilde{n}) \right)^2 + o_p(1).
\]

A Taylor expansion of \( \Phi^{-1} \) around \( i/\tilde{n} \) yields

\[
\Phi^{-1}(F_n^{-1}(i/\tilde{n})) - \Phi^{-1}(i/\tilde{n}) = \frac{1}{\phi(\Phi^{-1}(i/\tilde{n}))} (F_n^{-1}(i/\tilde{n}) - i/\tilde{n}) + R_n(i, n).
\]

The remainder term \( R_n(i, n) \) fulfills \( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} R_n(i, n) = o_p(1) \) because the empirical quantile process converges uniformly to a standard Brownian bridge \( B(\cdot) \), i.e. \( \sqrt{n}(F_n^{-1}(r) - r) \overset{d}{\rightarrow} B(r), r \in [0, 1] \), see e.g. Koenker and Xiao (2002). Then, we have that

\[
a_{1n,3} = \frac{1}{n^2} \sum_{i=1}^{n} \frac{1}{\phi^2(\Phi^{-1}(i/\tilde{n}))} \left( F_n^{-1}(i/\tilde{n}) - i/\tilde{n} \right)^2 + o_p(1)
\]

and \( a_{1n,3} \overset{p}{\rightarrow} 0 \) with another application of the Cauchy-Schwarz inequality. Then also \( a_{1n,3} \overset{p}{\rightarrow} 0 \).

Similarly,

\[
a_{2n} = \frac{1}{n} \sum_{i=1}^{n} z_i \eta_i + \frac{1}{n} \sum_{i=1}^{n} z_i (\hat{\eta}_i - \eta_i) =: a_{2n,1} + a_{2n,2}.
\]

Here, \( a_{2n,1} \overset{p}{\rightarrow} \mathbb{E}(z_i \eta_i) \) and \( a_{2n,2} \overset{p}{\rightarrow} 0 \) with the same reasoning as for \( a_{1n,2} \) so that \( a_{2n} \overset{p}{\rightarrow} \mathbb{E}(z_i \eta_i) \).

Also in the term \( b_n := n^{-1} \hat{\eta}' \hat{\eta}, \hat{\eta} \) can be replaced with \( \eta_i \) and it follows that \( b_n \overset{p}{\rightarrow} \mathbb{E}(\eta_i^2) = 1 \).

Summed up, the limit matrix \( S \) is given by

\[
S = Q - (0_k, \mathbb{E}(z_i \eta_i))'(0_k, \mathbb{E}(z_i \eta_i)).
\]

The positive definiteness of \( S \) follows from the identifying Assumption 2. \( \square \)
Proof of Lemma 3

Note that $\varepsilon$ is independent of all regressors. By the central limit theorem, $\frac{1}{\sqrt{n}} W' \varepsilon$ converges to a random variable distributed as $\mathcal{N}(0, \sigma^2 \epsilon Q)$. The second summand is equal to

$$\left( \frac{1}{n} W' \hat{\eta} \right) \left( \frac{1}{\sqrt{n}} \varepsilon \right) \left( \frac{1}{\sqrt{n}} \eta' \hat{\eta} \right)$$

The calculations from the proof of Lemma 2 yield that $\frac{1}{n} W' \eta \overset{p}{\to} (0_k, E(z_i \eta_i))'$ and $\frac{1}{n} \hat{\eta}' \hat{\eta} \overset{p}{\to} 1$. Moreover, $\frac{1}{\sqrt{n}} \varepsilon \eta' = \frac{1}{\sqrt{n}} \eta' \varepsilon + \frac{1}{\sqrt{n}} (\hat{\eta}' - \eta) \varepsilon$. The first term in brackets is asymptotically normally distributed with expectation zero and the second term vanishes asymptotically. To see this, note that

$$\frac{1}{\sqrt{n}} (\hat{\eta} - \eta)' \varepsilon = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varepsilon_i \left( \Phi^{-1}(F_n^{-1}(i/n)) - \Phi^{-1}(i/n) \right) + o_p(1),$$

as argued in the proof of Lemma 2. Results from stochastic integration yield $\frac{1}{\sqrt{n}} (\hat{\eta} - \eta)' \varepsilon \overset{p}{\to} 0$.

Let $\tilde{\varepsilon}_i = \frac{\varepsilon_i}{\sqrt{n}}$ represent an increment of a Brownian motion $W(\cdot)$ which is independent of $F_n(\cdot)$ such that $\sum_{i=1}^{[rn]} \tilde{\varepsilon}_i \Rightarrow W(r)$. It follows that

$$\sum_{i=1}^{n} \varepsilon_i \left( \Phi^{-1}(F_n^{-1}(i/n)) - \Phi^{-1}(i/n) \right) \overset{d}{=} \sum_{i=1}^{n} \varepsilon_i \sqrt{n} \left( \Phi^{-1}(F_n^{-1}(i/n)) - \Phi^{-1}(i/n) \right) \overset{d}{=} \int_0^1 B(r) dW(r) = O_p(1).$$

It follows that $\frac{1}{\sqrt{n}} (\hat{\eta} - \eta)' \varepsilon \overset{p}{\to} 0$. \hfill \Box

Proof of Lemma 4

It holds $G_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Phi^{-1}[F(\hat{e}_i)] \left( \Phi^{-1}(F(e_i)) - \Phi^{-1}[F(\hat{e}_i)] \right) + o_p(1)$. With the same rea-
soning as in the proofs of Lemma 2 and 3 we have
\[
G_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Phi^{-1}(i/\tilde{n}) \left( \Phi^{-1}(F_n^{-1}(i/\tilde{n})) - \Phi^{-1}(i/\tilde{n}) \right) + o_p(1)
\]
\[
= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\Phi^{-1}(i/\tilde{n})}{\phi(\Phi^{-1}(i/\tilde{n}))} (F_n^{-1}(i/\tilde{n}) - i/\tilde{n}) + o_p(1)
\]
\[
\xrightarrow{d} \int_0^1 \frac{\Phi^{-1}(r)}{\phi(\Phi^{-1}(r))} B(r) dr =: G.
\]

The convergence in the third row follows again from Koenker and Xiao (2002) and the continuous mapping theorem. Let \(g(r) := \Phi^{-1}(r)/\phi(\Phi^{-1}(r))\). Results collected in Webel and Wied (2016) yield that \(F\) is normally distributed with expectation zero and variance
\[
Var(G) = E \left( \left( \int_0^1 g(r) B(r) dr \right)^2 \right)
\]
\[
= \int_0^1 \int_0^1 g(s)g(u) E(B(s)B(u)) dsdu
\]
\[
= \int_0^1 \int_0^1 g(s)g(u)(\min(s,u) - su) dsdu.
\]
Numerical approximations then show that \(Var(G) = 1/2\). \(\square\)

Proof of Lemma 5

Let \(\tilde{x}_i = x_i - \mathbb{E}(x_i)\) and \(\tilde{w}_i = (1, \tilde{x}_i', z_i)'\) such that
\[
D = \begin{pmatrix}
1 & 0 & 0 \\
-\mathbb{E}(x_i) & I_{k-1} & 0 \\
0 & 0 & 1
\end{pmatrix}
\]
Under Assumption 11 it holds that
\[
\tilde{H}_n = \begin{pmatrix}
\tilde{H}_{n,1} \\
\vdots \\
\tilde{H}_{n,k+1}
\end{pmatrix} := D^{-1} H_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{w}_i \left( \Phi^{-1}(F_n^{-1}(i/\tilde{n})) - \Phi^{-1}(i/\tilde{n}) \right) + o_p(1),
\]
as argued in the proofs of Lemma 2, 3 and 4. The limiting distribution of each component 
$j = 1, \ldots, k + 1$ depends on the properties of the respective component in $\tilde{w}_i$. We need to 
consider three cases: (i) the intercept $\tilde{w}_{i,1} = 1$, (ii) $\tilde{w}_{i,j}$ with $j = 2, \ldots, k$ corresponds to the 
demeaned exogenous regressors. (iii) $\tilde{w}_{i,k+1}$ corresponds to the endogenous regressor.

(i) In this case,
$$
\tilde{H}_{n,1} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left( \Phi^{-1}(F_n^{-1}(i/\bar{n})) - \Phi^{-1}(i/\bar{n}) \right) + o_p(1),
$$
and arguments from the proof of Lemma 4 yield $H_{n,1} \overset{d}{\to} \int_{0}^{1} B(r)dr$ for a standard 
Brownian bridge $B$. This integral is normally distributed.

(ii) In this case for $j = 2, \ldots, k$ we have
$$
\tilde{H}_{n,j} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{x}_{i,j} \left( \Phi^{-1}(F_n^{-1}(i/\bar{n})) - \Phi^{-1}(i/\bar{n}) \right) + o_p(1),
$$
where $\tilde{x}_{i,j}$ has expectation zero and is independent of $(\Phi^{-1}(F_n^{-1}(i/\bar{n})) - \Phi^{-1}(i/\bar{n}))$. Results
from stochastic integration yield that $H_{n,j} \overset{p}{\to} 0$, as argued similarly in the proof of Lemma 3.

(iii) In this case,
$$
\tilde{H}_{n,k+1} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_i \left( \Phi^{-1}(F_z(z_i)) - \Phi^{-1}[\hat{F}_z(z_i)] \right) + o_p(1)
$$
$$
= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} F_z^{-1}(F_z(z_i)) \left( \Phi^{-1}(F_z(z_i)) - \Phi^{-1}[\hat{F}_z(z_i)] \right) + o_p(1)
$$
$$
= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} F_z^{-1}(F_n^{-1}(i/\bar{n})) \left( \Phi^{-1}(F_n^{-1}(i/\bar{n})) - \Phi^{-1}(i/\bar{n}) \right) + o_p(1).
$$
Similarly as in the proof of Lemma 4
$$
H_{n,j} \overset{d}{\to} \int_{0}^{1} \frac{F_z^{-1}(r)}{\phi(\Phi^{-1}(r))} B(r)dr =: H_j
$$
. Let $h_j(r) := \frac{F_z^{-1}(r)}{\phi(\Phi^{-1}(r))}$. Then $H_j$ is normally distributed with expectation zero and variance
$$
Var(H_j) = \int_{0}^{1} \int_{0}^{1} h_j(s)h_j(u)(\min(s,u) - su)dsdu. \tag{A.2}
$$
The lemma holds for the case that the variance (A.2) exists and is finite. □
Proof of Theorem 1

We first prove part b). From the proof of Lemma 2 we know that

\[
\frac{1}{n} W' \hat{\eta} \xrightarrow{p} \begin{pmatrix} 0_k \\ \mathbb{E}(z_i \eta_i) \end{pmatrix}.
\]

Accordingly, with the results of the previous lemmas we get

\[
\sqrt{n} (\alpha - \alpha) = \sqrt{n}(W'M_\hat{\eta}W)^{-1}W'M_\hat{\eta}\varepsilon + \sqrt{n}(W'M_\hat{\eta}W)^{-1}W'M_\hat{\eta}(\eta - \hat{\eta})\rho
\]

\[
= \left( \frac{1}{n} W'M_\theta W \right)^{-1} \left( \frac{1}{n} W'M_\hat{\eta}\varepsilon + \frac{1}{n} W'(\eta - \hat{\eta}) \rho \right)
\]

Asymptotic normality then follows from the joint convergence (Assumption 1.1).

For proving part a), note that similarly as in the proof of part b),

\[
\hat{\alpha} - \alpha = \left( \frac{1}{n} W'M_\theta W \right)^{-1} \left( \frac{1}{n} W'M_\hat{\eta}\varepsilon + \frac{1}{n} W'(\eta - \hat{\eta}) \rho \right)
\]

The term \( \frac{1}{n} W'(\eta - \hat{\eta}) \) converges to zero in probability with the same argument based on the Cauchy-Schwarz inequality as used for the proof of Lemma 2. Assumption 2 is not necessary for this.

Proof of Theorem 2

Under the null hypothesis the \( t \)-statistic can be written as

\[
\frac{\hat{\eta}'\varepsilon - \hat{\eta}'W(W'W)^{-1}W'\varepsilon}{\hat{\sigma}_\varepsilon \sqrt{\hat{\eta}'\hat{\eta} - \hat{\eta}'W(W'W)^{-1}W'\hat{\eta}}}
\]

where \( \hat{\sigma}_\varepsilon^2 \) is the usual estimator of the residual variance. Using Theorem 1 it is not difficult to show that \( \hat{\sigma}_\varepsilon^2 \xrightarrow{p} \sigma_\varepsilon^2 \). Furthermore, using the results of Lemma 2 we have \( n^{-1}\hat{\eta}'\hat{\eta} = \ldots \)
\[ n^{-1} \eta' \eta + o_p(1) \xrightarrow{p} 1 \text{ and} \]
\[ \frac{1}{n} \eta' W(W'W)^{-1} W' \hat{\eta} = \frac{1}{n} \eta' W(W'W)^{-1} W' \eta + o_p(1) \]

For the numerator of the \( t \)-statistic we use the results of Lemma 3 and obtain
\[ \frac{1}{\sqrt{n}} \hat{\eta}' \varepsilon = \frac{1}{\sqrt{n}} \eta' \varepsilon + o_p(1) \]
\[ \xrightarrow{d} N(0, \sigma^2_\varepsilon) \]

In Lemma 2 we show that
\[ \frac{1}{n} \hat{\eta}' W = \frac{1}{n} \eta' W + o_p(1). \]

Collecting these results allows for representing the \( t \)-statistic as
\[ \frac{\eta'(I_n - W(W'W)^{-1} W') \varepsilon}{\sigma_\varepsilon \sqrt{\eta'(I_n - W(W'W)^{-1} W') \eta}} + o_p(1). \]

Accordingly, under the null hypothesis \( \rho = 0 \) the estimation error \( \hat{\eta} - \eta \) does not enter the asymptotic distribution and the test statistic is asymptotically equivalent to the test statistic when \( \eta \) is known. From standard results for the linear regression it immediately follows that the test statistic has the usual standard normal limiting distribution. \( \square \)