Can direct payments facilitate agricultural commercialisation: Evidence from a transition country

Philip Kostov, Sophia Davidova, Alastair Bailey, Ekrem Gjokaj and Kaplan Halimi

On-line Appendices.

Appendix 1. Naïve regression estimate

Coefficients	Estimate	P-Value
(Intercept)	12.17	0.04
sub_grain	0.01	0.08
sub_fr_veg	0.04	0.00
sub_liv	0.01	0.06
V14_1_i_market_km	0.43	0.15
V14_1_k_munic_km	-0.85	0.01
risk_att	2.84	0.00
V12_01_eduyears_resp	0.22	0.54
V09_1_useany	8.01	0.09
V09_1_a_use_fert	-5.28	0.18
advisory	8.09	0.01

Appendix 2. Naïve conventional instruments estimation

Coefficients	Estimate	P-Value
(Intercept)	90.20	0.31
sub_fr_veg	-0.56	0.42
sub_grain	0.42	0.48
sub_liv	-0.31	0.55
V14_1_k_munic_km	-0.86	0.28
risk_att	-0.62	0.90
V12_01_eduyears_resp	1.37	0.38
V09_1_useany	-3.20	0.83
V09_1_a_use_fert	0.16	0.99
advisory	0.60	0.97
HH_labour_only	-63.45	0.37

	Statistic	P-value
Weak Instruments tests:		
(sub_fr_veg)	1.37	0.25
(sub_grain)	1.58	0.18
(sub_liv)	0.75	0.56
Wu-Hausman	1.20	0.31
Hansen J-test	0.01	0.95

Appendix 3 Copula estimation results

Coefficients	Estimate	P-value
Intercept	13.34	0.79
V14_1_k_munic_km	-0.68	0.04
risk_att	3.14	0.00
V12_01_eduyears_resp	0.21	0.69
V09_1_useany	3.56	0.59
V09_1_a_use_fert	-7.50	0.16
advisory	1.80	0.67
sub_fr_veg	0.07	0.04
HH_labour_only	-15.31	0.04
sub_liv	0.00	0.98
sub_grain	-0.06	0.39
PStar.sub_fr_veg	-9.55	0.53
PStar.sub_liv	8.48	0.58
PStar.sub_grain	14.76	0.46

Appendix 4. Lewbel estimation

	Estimate	P-Value
sub_fr_veg	0.04	0.03
sub_grain	-0.10	0.50
sub_liv	-0.09	0.27
(Intercept)	31.17	0.03
V14_1_k_munic_km	-0.75	0.02
risk_att	3.71	0.01
V12_01_eduyears_resp	0.25	0.64
V09_1_useany	1.91	0.80
V09_1_a_use_fert	-5.28	0.41
advisory	4.11	0.45
HH_labour_only	-13.09	0.15
	J-test	P-value
Test E(g)=0:	4.41	0.73

Partial F-test for weak identification:		
sub_fr_veg	13.38	
sub_grain	1.15	
sub_liv	2.71	

Appendix 5. Partial identification models

A. Fruits and vegetables DP

Coefficients	Estimate	P-value
Intercept	34.51	0.00
V14_1_k_munic_km	-0.67	0.04
risk_att	3.08	0.00
V12_01_eduyears_resp	0.22	0.67
V09_1_useany	3.58	0.58
V09_1_a_use_fert	-7.17	0.18
advisory	1.77	0.67
HH_labour_only	-15.92	0.04
sub_fr_veg	0.04	0.04

	Statistic	P-value
Weak instruments	193.50	0.00
Wu-Hausman	3.32	0.07
Sargan	3.12	0.68

B. Grains and oilseeds DP

Coefficients	Estimate	P-value
Intercept	42.66	0.00
V14_1_k_munic_km	-0.70	0.05
risk_att	2.13	0.09
V12_01_eduyears_resp	0.24	0.67
V09_1_useany	3.06	0.67
V09_1_a_use_fert	-7.74	0.19
advisory	-0.39	0.93
HH_labour_only	-20.52	0.01
sub_grain	0.12	0.03

	Statistic	P-value
Weak instruments	7.39	0.00
Wu-Hausman	8.29	0.00
Sargan	3.43	0.63

C. Livestock DP

Coefficients	Estimate	P-value
Intercept	37.90	0.00
V14_1_k_munic_km	-0.70	0.04
risk_att	3.09	0.01
V12_01_eduyears_resp	0.31	0.57
V09_1_useany	1.74	0.80
V09_1_a_use_fert	-5.98	0.28
advisory	2.32	0.60
HH_labour_only	-18.08	0.02
sub_liv	-0.11	0.02

	Statistic	P-value
Weak instruments	49.27	0.00
Wu-Hausman	15.88	0.00
Sargan	2.29	0.68

Appendix 6. Estimation results for Zeros inflated beta regression models with different mean ling=k function and explanatory variable for variance ands inflation equation

	logit		probit		log		cloglog	
Mean	Estimate	P value	Estimate	P value	Estimate	P value	Estimate	P value
Intercept	-0.5259	0.0459	-0.3265	0.0457	-0.9613	0.0000	-0.7533	0.0002
V14_1_k_munic_km	-0.0209	0.0377	-0.0130	0.0368	-0.0104	0.0514	-0.0151	0.0503
risk_att	0.0856	0.0124	0.0535	0.0120	0.0418	0.0142	0.0607	0.0184
V12_01_eduyears_res p	0.0092	0.5372	0.0056	0.5462	0.0048	0.5528	0.0073	0.5194
V09_1_useany	0.3890	0.0720	0.2414	0.0723	0.1841	0.0767	0.2802	0.0733
V09_1_a_use_fert	-0.4517	0.0106	-0.2806	0.0106	-0.2088	0.0067	-0.3186	0.0097
advisory	0.3481	0.0047	0.2185	0.0045	0.1685	0.0036	0.2425	0.0064
HH_labour_only	-0.1350	0.1602	-0.0843	0.1594	-0.0773	0.1280	-0.1024	0.1563
sub_fr_veg	0.0014	0.0000	0.0008	0.0000	0.0004	0.0000	0.0008	0.0000
sub_grain	0.0005	0.0000	0.0003	0.0000	0.0002	0.0000	0.0003	0.0000
sub_liv	0.0001	0.0000	0.0001	0.0000	0.0001	0.0000	0.0001	0.0000
v1	-0.0019	0.0013	-0.0011	0.0021	-0.0005	0.0219	-0.0011	0.0038
v2	-0.0008	0.0057	-0.0005	0.0064	-0.0003	0.0460	-0.0005	0.0154
v3	-0.0003	0.0000	-0.0002	0.0000	-0.0001	0.0000	-0.0002	0.0000
Variance								
Intercept	0.6830	0.0000	0.6843	0.0000	0.7144	0.0000	0.6965	0.0000
sub_fr_veg	0.0021	0.0086	0.0020	0.0119	0.0014	0.0038	0.0019	0.0164
sub_grain	0.0011	0.0016	0.0011	0.0019	0.0009	0.0005	0.0010	0.0033
sub_liv	0.0027	0.0000	0.0027	0.0000	0.0026	0.0000	0.0027	0.0000
V14_1_k_munic_km	0.0263	0.0504	0.0264	0.0500	0.0245	0.0539	0.0254	0.0590
Zeros Inflation								
Intercept	0.1552	0.5267	0.3971	0.1336	0.3971	0.1334	0.3971	0.1334
sub_fr_veg	-0.0098	0.0465	-0.0097	0.0459	-0.0097	0.0459	-0.0097	0.0459
sub_grain	-0.0012	0.0658						
sub_liv	-0.0048	0.0640	-0.0050	0.0576	-0.0050	0.0576	-0.0050	0.0576
risk_att	-0.1408	0.0061	-0.1236	0.0176	-0.1236	0.0175	-0.1236	0.0175
V09_1_a_use_fert			-0.5060	0.0070	-0.5060	0.0070	-0.5060	0.0070

Appendix 7

Technical implementation details

The paper implements a range of non-standard estimation techniques. This appendix provides details of their implementation that can potentially benefit interested researchers.

Copula based endogeneity correction (Park, and Gupta, 2012)

The empirical implementation in this paper deals with a particularly straightforward application of the above method, namely several endogenous variable, all of which are continuous. This means that the implementation uses the control function type of transformation for the endogenous variables to augment the model by additional auxiliary variables. Technically this involves the following steps for each endogenous variable:

a) estimate the empirical distribution function for the variable

b) correct the margins (i.e. transform values of 1 and 0 , so the next step can be applied)

c) using the assumed marginal distribution (in this case normal) estimate the s normal quantile of the values obtained in the previous step

For each continuous endogenous variables these steps create an auxiliary variable to be added alongside the original endogenous variable to the regression model.

Technically the steps above are very simple statistical calculations and can be carried out by any statistical software package (even Excel). The only complication is that since this results in a two step procedure, valid inference requires bootstrapping the whole procedure (auxiliary variables creation and following regression). In practical terms this means that instead of simply transforming the endogenous variables we need a simple function that is called within each bootstrap iteration.

A typical implementation of such a function will contain 4 lines of code (step b needs 2)

See for example the R function below

```
get.pstar <- function(x){

mp <- ecdf(x)

mp[mp == 0] <- 10^(-7)

mp[mp == 1] <- 1-10^(-7)

qnorm(mp)

}
```

The R statistical system, contains a contributed package REndo, which contains a ready made function for copula correction estimation, that has been used for the results in Appendix 3.

To obtain copula correction estimates corrected for multicollinearity (such as the ridge estimates in Table 3) we used customised code which relied upon the R packages boot (for bootstrap, not essential, can be made manually instead), lmridge (for ridge estimation and implementation of various penalty selection criteria) and pls and plsRglm (for partial least squares estimation). We did test for multicollinearity using the mctest package (results not included in the paper itself)

R offers plenty of alternative options for both ridge regression and PLS.

The Lewbel (2012) approach is available in both the ivlewbel and REndo packages. The former is no longer supported but is easier to use and more flexible. REndo only allows a single endogenous variable, so cannot be used for the purposes of this paper .

The Lewbel (1997) suggestions (with a limited number of non-linear transformation functions) are implemented in Rendo. Strictly speaking REndo can only handle a single endogenous variable.   Creating customised routines to apply the relevant transformed 'instruments' is not technically difficult (although labour intensive). One can (as we have done) modify some of the REndo source code to do this. Using this we can calculate the relevant instruments for the multiple endogenous variables to use in standard instrumental variable routine (i.e. in the ivreg function in the AER package) to estimate the first model in Table 5 or the corresponding moment conditions for use with the gmm package (as in the second model in Table 5, and a range of similar unreported alternative estimates).

Finally for fractional response models. The relevant instruments were calculated by using the customised functions as per above.

The GMMz estimator of Ramalho and Ramalho (2017) is available in the frhet package.

Both the zeros inflated Beta and Poisson regression models were estimated using the gamlss package, mainly because of its flexibility and ease of use, but there are a number of alternatives.

In particular the gamlss allows one to use user-defined distribution functions and also contains convenience functions that allow one produce inflated versions of any distributions providing extra flexibility. Bayesian equivalents for the gamlss model can be estimated via the bamlss package. The latter package is useful in that in addition to Bayesian (and frequentist) versions of such models contains a large number of utility functions for boosting and bootstrap based estimation and model selection. One needs to be warned though that many of the extensive useful features of both these packages are not well documented.

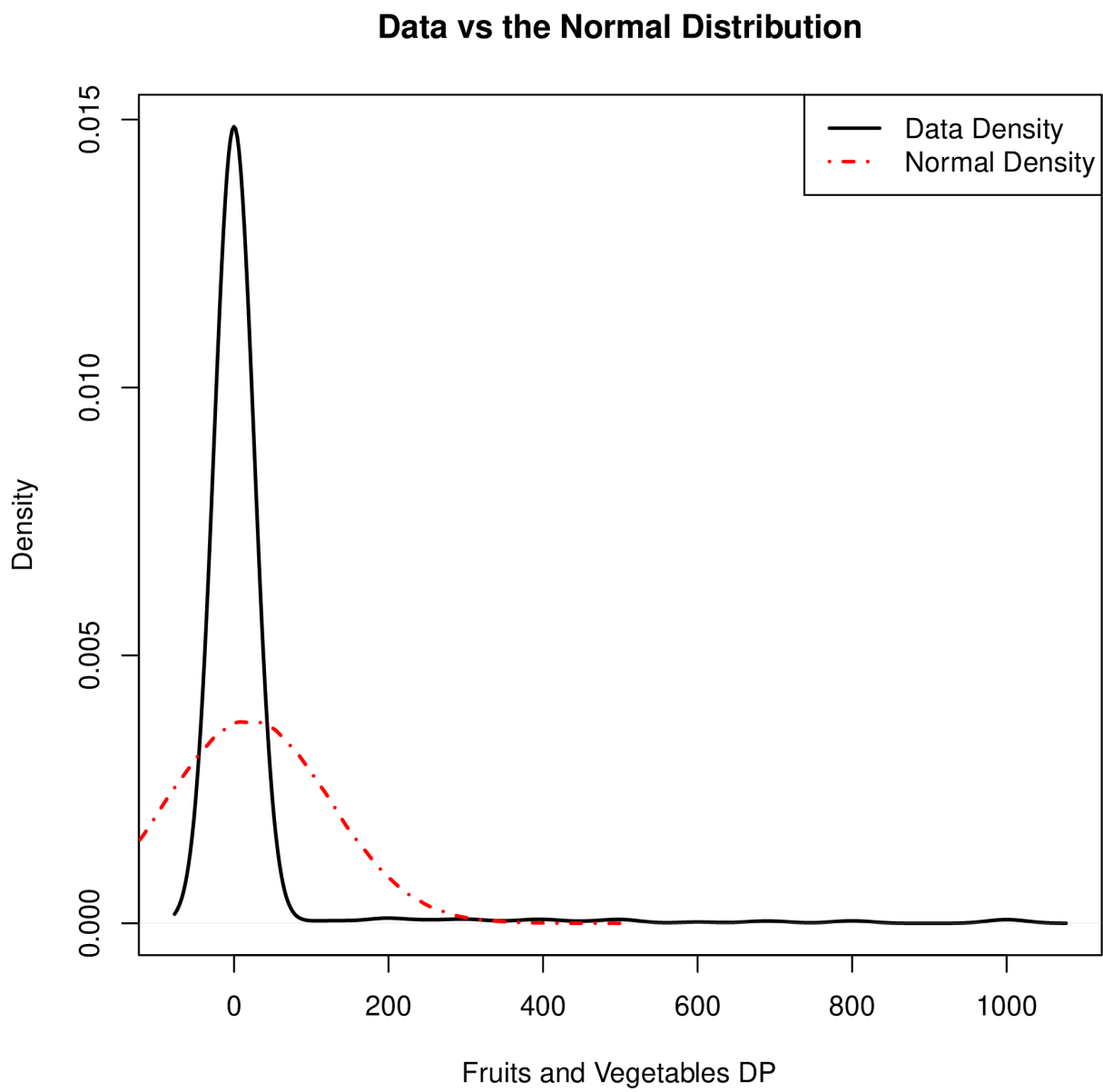Figure 1. Empirical density of the fruit and vegetable DPs

Data vs the Normal Distribution

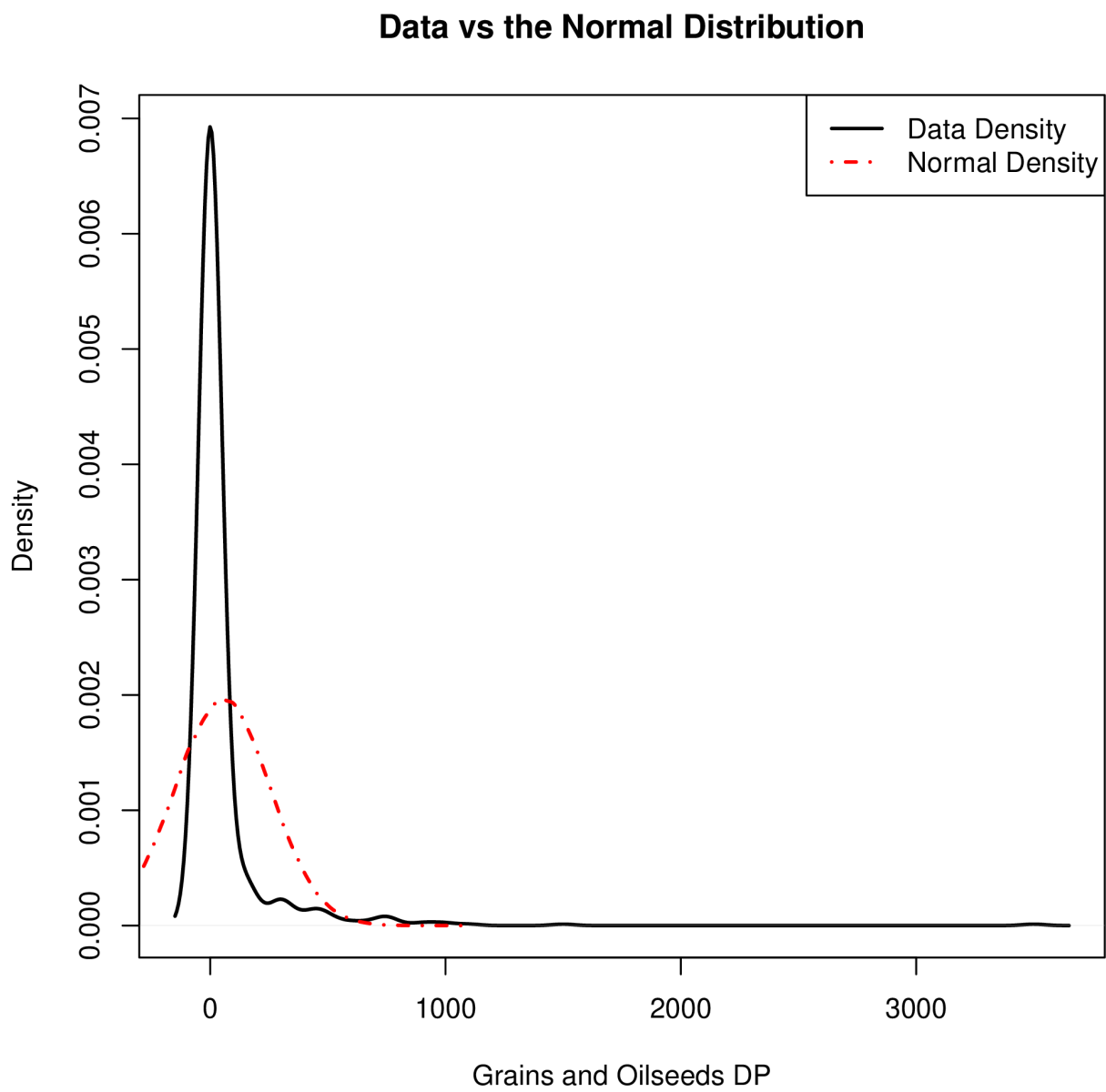Figure 2. Empirical density of the cereal and oilseed DPs

Data vs the Normal Distribution

Figure 3. Empirical density of the livestock DPs

Data vs the Normal Distribution