Edinburgh School of Economics

Discussion Paper Series

Number 321

OLS Estimation when Two Noisy Measures of a Regressor are Available: Instruments, Tests and Pitfalls

Andy Snell
The University of Edinburgh

June 2025

Published by
School of Economics, University of Edinburgh
30 -31 Buccleuch Place
Edinburgh, EH8 9JT
@econedinburgh

THE UNIVERSITY *of* EDINBURGH
School of Economics

OLS Estimation when Two Noisy Measures of a Regressor are Available: Instruments, Tests and Pitfalls

Andy Snell *(University of Edinburgh)*

June 2025

Summary

It is well known that if the regression coefficient of y on w (β say) has a constant probability limit but we only have two noisy measures of w - x and z - then we may obtain consistent estimates of β as long as a) the measurement errors are classical and b) the measurement errors are uncorrelated. We propose a simple test of a) and a test for b) as part of a composite null. To effect the latter we instrument x with z and functions of z and vice versa to obtain two sets of overidentifying restrictions tested via a standard J test of instrument validity. If no test in this sequence rejects we then combine the orthogonality conditions to obtain a single efficient estimate of β. We discuss the likely prior validity of the various instruments and the pitfalls in using the test procedure. Unlike standard overidentification tests which diverge in heterogeneous response settings even when each instrument is valid, our tests only diverge when one or more instrument is invalid. We apply the test sequence and estimation procedure to analyse i) the cyclical component of wages and ii) the effect of state level unemployment on burglaries in the US. Correcting for measurement error raises the estimates of β in both applications.

JEL Codes:

Keywords: Measurement Error, Instrumental Variables, Consistent OLS estimation.

[0]Email addresses: Snell: a.j.snell@ed.ac.uk;

1 Introduction and Overview

Regressor measurement error (ME) is a pervasive problem in survey data. There is accordingly a huge literature assessing its impact and - under suitable assumptions - proposing estimators that circumvent the problem in parametric and nonparametric settings. (For an excellent survey and overview of the literature see Bound et al. (2001) or more recently for a review focusing on nonlinear models see Chen et al. (2011)). In this paper we focus on a single thread of this problem namely the case where the sole object of interest is the least squares estimate of y on an unobserved regressor w (β) and where two noisy measures of w - x and z - are available. We propose a simple sequence of tests whose composite null is that the instruments are valid - a condition that ensures the estimate of β is consistent. The first test in the sequence has a null that ME's are classical in nature. The subsequent tests examine instrument validity. Use of instruments is standard in solving measurement error problems. However the novelty here is that we use functions of x to instrument z and functions of z to instrument x *together in a single procedure*. A key necessity is that β is the sole focus of interest. We now expand on the procedure and outline the paper's contents.

It is well known that if a) x and z are two noisy measures of an unobserved regressor w, b) the object of interest is an OLS estimate (β) of y on w, c) if the latter tends to a constant limit, d) the two measurement errors (henceforth ME's) are classical in nature and e) the ME's are uncorrelated, then using x as an instrument for z and/or vice versa will deliver consistent estimates of β. In this paper we take b) as given and adopt c) as an assumption - an assumption which is arguably weak. We propose a simple test to examine d) - the first test in a sequence of three. Condition e) however is more problematic. As Bound et al. (2001) point out, there is no direct way of testing the uncorrelatedness of ME's without a validation study. In particular and as we show below, basing a test on the difference between the estimate generated by using x as an instrument for z and z as an instrument for x is doomed to fail because under the alternative where the ME's are correlated both estimates have the same asymptotic bias. The second and third tests in our sequence overcome this problem. They separately test overidentification restrictions generated by using instruments based on x for z and vice versa.[1] In this scenario uncorrelatedness of ME's is part of a composite null that the instruments are

[1] If there are only sufficient relevant instruments to conduct one of the two overidentification tests then the sequence would be reduced to two tests. Of course if it is not possible to conduct either of the overidentification tests then we could not proceed at all. However even in that case the test for classical ME (the first in our sequence) is always executable and if it passes may allow us to bound the bias.

valid - a null that guarantees consistency of β. The instruments we propose are powers and (where available) lagged values of x for z and vice versa. Finally if each of the tests in the sequence fail to reject, we propose that estimation of β proceed by combining the orthogonality conditions underpinning the second and third tests. This last step would yield a further overidentification test. However, as we show below, this last test may suffer the pitfall alluded to above where despite two instruments being invalid each delivers the same biased estimate of β - a consequence that leads to test inconsistency.

We emphasise at this point that in a heterogeneous response setting, traditional overidentification tests of instrument validity diverge even if the instruments are valid. This is because in such settings, different valid instruments estimate different weighted averages of the heterogeneous responses of individual units (see for example Angrist and Imbens (2000)). By contrast and as we illustrate below, in our framework *all valid* instruments consistently estimate the same OLS parameter. [2] Another point worthy of emphasis is that our view of ME is very broad. We define it as any discrepancy between the desired regressor and the measured regressor. In the context of microeconometric survey data ME is clearly understood as a misreported response. However in a macroeconometric context ME could arise because a macro variable is estimated from micro level data. In one of our empirical applications we use estimates from the BLS of annual state level unemployment rates. As these are compiled from the CPS they may contain both sampling error (CPS samples are quite small) *and* misreporting error. Typically this scenario arises when macroeconomic stylised facts - rather than microeconometric causal effects - are the objects of interest.

In the next section we outline our proposed testing and estimation framework. In section 3 we examine power and size of the test sequence via a small simulation study. In section 4 we apply the tests and IV estimation in two scenarios:- a) the cyclicality of US real wages and b) the effects of local (state level) unemployment on burglaries in the US. The Local Area Unemployment Statistics(LAUS) used in applications such as b) have a high degree of ME as they are derived from small subsamples of various surveys, notably and mainly the CPS. We obtain two measures of the state unemployment rate by random sampling across counties. Using these two measures and controlling for ME raises estimates of the impact of local unemployment on crime substantially. In a) we treat the

[2] The quasi experimental framework of Angrist and Imbens treats heterogeneous responses as fixed parameters attached to individuals rather than as random variables. However Imbens and Rubens (2015) propose a Bayesian extension to the framework which envisages the treatment effects as being drawn from a meta distribution or "super population". Adopting this approach here allows our framework to accommodate heterogeneous responses of y to w.

issue of"true" cyclical measure as a latent variable problem. There we find that whilst the two individual measures of the business cycle deliver rather different OLS estimates of wage cyclicality, both are lower and less significant than that derived using our procedure.

2 Model and Assumptions

Consider the three equation system

$$y_i \equiv \beta_N w_i + \varepsilon_i^N ....i = 1, ..N \tag{1}$$

$$\sum_{i=1}^N w_i \varepsilon_i^N = 0 \tag{2}$$

$$x_i = w_i + u_i \tag{3}$$

$$z_i = w_i + v_i \tag{4}$$

$$y_i = \beta_N x_i + (\varepsilon_i^N - \beta u_i) \tag{5}$$

$$y_i = \beta_N z_i + (\varepsilon_i^N - \beta v_i) \tag{6}$$

$$|\rho_{xz}| < 1 \tag{7}$$

where w_i is an unobserved variable for which there are two noisy measures x_i and z_i and β_N is the OLS coefficient of y_i on w_i in a sample of size N. Equations (1) to (6) are purely definitional. Equation (1) simply defines the object of interest - the OLS regression of y on w (or more formally the linear projection of y on w). The assumption in (7) requires ME to exist and for the two measures to have some independent variation. Later on we will be instrumenting z with functions of x and vice versa in a single procedure. This will not be feasible if x and z are identical. We view this assumption as easy to check and rather weak; it merely defines what is meant when we say that two noisy measures of w are available. In what follows we drop the i subscript when discussing variables in the text wherever possible and $\sum$ will denote a summation over $i = 1, ..N$. In the Annex we show how the analysis is easily extended to allow for controls in (1).

We confine the analysis to settings in which $plim\{\beta_N\} = \beta$ i.e. where β_N has a constant probability limit. The actual interpretation of β will depend on the underlying model for the variables and will be context specific. Below we discuss a few examples including one where we may interpret β in terms of underlying heterogeneous responses of y to w. We wish to estimate β via GMM [3] using x and functions thereof as instruments

[3] Although we refer to β as a large sample OLS estimate more formally it is the parameter in the

for z and vice versa *in a single estimation procedure*. Our sole focus on β means we are effectively estimating a linear model. Although the assumptions required for GMM to be a valid tool for inference in this context are standard[4], our "reflexive" use of instruments is not. In particular we need to be sure that the the orthogonality conditions we use have a nonsingular covariance matrix. In the Annex we give sufficient conditions for this to hold and show that the use of "reflexive" instruments itself does not generically lead to the failure of this assumption. Finally our use of powers of x and z as instruments requires the existence of moments of higher order than is typically required (up to order eight in our case). The test of Trapani (2016) could be used to establish the existence of such moments.

2.1 Heterogeneous Responses

The assumption that $plim\{\beta_N\} = \beta$ is not trivial in settings of heterogeneous responses. For example suppose that y_i was equal to $f(w_i, \gamma_i, \sigma_i \xi_i)$ where ξ_i is an *iid* random (unobserved) error term independent of w_i and γ_i a vector of parameters defining the response of y_i to w_i. In this scenario we would require inter alia that (σ_i, γ_i) be drawn from a pdf with a finite number of fixed parameters. (See for example Pesaran (2015) who analyses the case where f is linear, w is exogenous and γ_i are heterogenous random slopes. See also Imbens and Rubens (2015) who allow for treatment effects to be drawn from a "super population".) To take a more specific case consider a random parameter model with heterogenous slopes β_i.

$$y_i = \beta_i w_i + \upsilon_i \tag{8}$$

$$E(\beta_i) = \overline{\beta} \quad where \tag{9}$$

$$y_i = \overline{\beta} w_i + \varsigma_i \tag{10}$$

$$\varsigma_i = \upsilon_i + (\beta_i - \overline{\beta}) w_i \tag{11}$$

$$= \upsilon_i + \xi_i^{\beta} w_i \tag{12}$$

asymptotic linear projection of y on w. It can be estimated in a number of ways including GMM

[4] Hansen (1982) seminal paper gives general conditions for GMM to offer valid estimation and inference for time series data and in a more recent paper Kuersteiner and Prucher(2013) gives conditions for panel data where "T" is fixed but "N'' is large.

[5] Consider the case where w is uncorrelated with both $\xi^{\beta}w$ and ν. In this scenario $\beta = \overline{\beta}$ and the large sample OLS estimate of y on w (β) has a clear causal interpretation. By contrast if $\xi^{\beta}w$ was correlated with w then $\beta = \overline{\beta} + \frac{\sigma_{ww\xi\beta}}{\sigma_{w^2}}$ where here and henceforth σ_a denotes $plim(\sum_{i=1}^{N} a_i/N)$. Alternatively if w was not strictly exogenous and no causal interpretation was possible β would be a data moment. Such a quantity would be useful when calibrating a theoretical model in which w and y were jointly determined and where a data value for β was required as a target. However we re-emphasise that this paper addresses the impact of ME on the large sample estimate of β and so the interpretation of β is not central to it.

The overidentification tests we propose are analagous to those traditionally used to examine instrument validity in the fixed response model without ME. As is well known, in a model with heterogeneous responses, overidentification tests may reject the null even when instruments are valid - valid in the sense of being uncorrelated with the equation's error term. This issue does not arise in our framework as we now explain. Consider again equation (8) above but now where w is thought to be correlated with ν. The investigator may have two instruments at her disposal and wish to test their validity in the traditional way. Both instruments may well be uncorrelated with υ and hence be valid in the traditional sense. However if their correlation with β_i (equivalently with ξ_i^{β}) differs then the overidentification test of their validity will diverge. In our framework things are different. To illustrate, reconsider the case of (10) but now with w being uncorrelated with the composite error term ς. The OLS estimate of y on w here consistently estimates β. We do not observe w so treating z (or x) as the regressor instead we propose to use x (or z) and functions of x (or z) to instrument for z (or x). We would then test the overidentification restrictions arising from having more than one instrument using a standard J test. In this scenario a key part of our null hypothesis is that the ME's are uncorrelated with ς_i. In turn this would require that the ME's be uncorrelated with the random response parameter β_i. For us this is what instrument validity means. A rejection of our null does not speak to instrument validity in the traditional sense therefore. Instead it merely tells us that IV does not deliver consistent estimates of the OLS parameter β - something that is our sole focus of interest. [6]

Later we will be seeking two sets of instruments. One set will be arguably uncor-

[5]This issue is resonant of the debate between Heckman (1999) There Heckman argues that the IV estimate only converges to the treatment on the treated parameter if the size of response - here ξ^{β}- does not cause agents to select into treatment

[6]There are cases where failure of parts of the composite null to hold do not undermine consistency. We discuss this below and argue that such occurrences are likely to be relatively rare in practice.

related with $y - \beta x$. and the other arguably uncorrelated with $y - \beta z$. This will yield two sets of overidentifying restrictions. Assuming first lags (denoted by x_{-1} and z_{-1}) are available, the instruments we propose to use are, respectively, x, x^3, and x_{-1} for z and z, z^3 and z_{-1} for x, choices we defend below.

3 Existing Literature

Perhaps the closest paper to ours in terms of the specifics of what it sets out to achieve is that of Wilhelm (2019) who proposes a test for the existence of ME. Like us his sole focus is consistent estimation of a single function of the data. However our starting position is that ME *does* exist (our procedure would collapse if this were not true) and is non trivial in terms of its impact on estimation. For survey data this is likely to be a very sound pretext. Additionally the approach in Wilhelm is non parametric whereas our approach is simpler and likely to be more familiar to empirical economists. Having said this if there is doubt as to whether or not ME has a substantial impact on least squares estimates then Wilhelm's test could certainly be used in the first instance. A rejection would naturally require a solution to the problem such as the one we propose here. Another paper tangentially related is that of Anderssen and Moen (2015). Under the assumption of uncorrelated ME's they show how to optimally combine the two arguably consistent estimates of β obtained from using z as an instrument for x and vice versa. In our paper the crucial assumption that ME's are uncorrelated is tested for not assumed a priori and we use the GMM apparatus to obtain a single estimate of β when it is overidentified.

The use of functions of measures as instruments to surmount ME issues is of course not new. In particular, Dagenais and Dagenais (1997) and Lewbel (1997) propose the use of higher order terms in regressors as instruments. Both make strong assumptions on the nature of ME inter alia requiring them to be classical. On the other hand their approach does not require the existence of two measures. In a similar vein Erickson and Whited (2002) use transformations of the regressor as instruments but they too require very strong assumptions on ME - inter alia that the ME's are classical in nature and uncorrelated with each other. Unlike these papers we offer tests of the key assumptions validating our procedure. A further novelty here is the *simultaneous* use of the two measures to instrument each other in a "reflexive" manner.

We require inter alia that ME is classical in nature and the first test in our sequence examines this. Bound et al. (2001) discuss the results of a number of ME validation studies which tend to suggest ME is not classical. We do not take a prior position on whether or not these results generalise to wider contexts. We would hope that if ME was non classical in any particular application then our test would signal this via a rejection. Of course a failure to reject does not necessarily imply the absolute "truth" of the null. It could also be consistent with a violation of the null that is too quantitatively small to be of consequence.

4 Consistent Estimation of the OLS Parameter

In what follows we ignore controls in (1) and WLOG take all variables to have a mean of zero. Adding controls does not - under reasonable assumptions - change what we present here as we show in the annex. A baseline assumption in GMM is that all variables are ergodic so that $E(a)= plim(\sum_{i=1}^{N} a_i/N)= \sigma_a$. Using this assumption we can couch the discussion in terms of probability limits rather than expectated values.

We begin by re-examining the idea that consistent estimates of β may be obtained by using x as an instrument for z (or vice versa) if ME's are classical and uncorrelated. A test of a composite null that implies the former is straightforward. Consider the OLS regression

$$x_{ij} - z_{ij} = u_{ij} - v_{ij} = \gamma y_{ij} + \epsilon_{ij} \tag{13}$$

A t-test of $\gamma = 0$ [7] (henceforth referred to as "T_C") is in effect a test of of a). More properly, the null that w and ε are uncorrelated with both u and v implies

$$\gamma\{ = \frac{\sigma_{xy} - \sigma_{zy}}{\sigma_{yy}}\} \tag{14}$$

$$= \frac{1}{\sigma_{yy}}\{\beta(\sigma_{wu} - \sigma_{wv}) + (\sigma_{u\varepsilon} - \sigma_{v\varepsilon})\} = 0 \tag{15}$$

This is a useful test but has a caveat. If x and z have a common additive element that is non classical then differencing will wash that element out. If additionally the non common components are classical then (large sample) power will equal size and the test will be inconsistent. Whilst the idea that the null will fail purely because of common components is rather pathological, the removal of such components via differencing may

[7] The investigator may cluster standard errors along dimensions that fit her priors about the nature of ME correlation..

reduce test power. [8] Notwithstanding this caveat T_C is a useful test of classical ME and whilst here we envisage using it as part of a test sequence it also stands alone as a device to offer inference on the nature of the ME's. Finally, as an alternative, we could use a Fisher non parametric test (Fisher (1922)) to test independence of $(x-z)$ and y instead. Independence is stronger than the required zero covariance and the test would still suffer the common elements issue highlighted above. However, many of the restrictions in the composite null hypothesese we consider below are automatically satisfied if independence holds so passing a Fisher test may well add power to our overall testing procedure.

Interestingly (14) shows T_C to be the scaled up difference between two IV estimates - one using x as an instrument for z and vice versa. A priori we might view such a test as shining a light on the consistency of the two estimates; if they differ they cannot both be consistent but if they do not differ they will be. However as we have discussed already if σ_{uv} is non zero then both estimates are biased, a bias that as (14) shows cancels out of the difference. Our strategy in this paper therefore is to augment T_C with two other tests that have uncorrelatedness of ME's as part of a composite null. The two tests are generated, respectively, by two sets of orthogonality conditions; one where instruments based on x are used (for z) and the other where instruments based on z are used (for x). We call these tests T_x and T_z. We test the overidentifying restrictions in each case. If both tests fail to reject, we would then proceed to combine the two sets of instruments into a single GMM procedure yielding a single estimate of β.

Here we examine the prior credibility of the overidentifying restrictions implied by the validity of each instrument in the two groups. To do this we derive an expression for the large sample estimate of β delivered by each instrument to expose the corresponding restrictions on the underlying data moments that ensure consistency. We also need to check that if instruments in a group *are* invalid they do not lead to the same biased estimates of β. Were that to be the case an overidentification test would fail to diverge and give misleading inference[9].
Choosing appropriate instruments in any application is always a matter for heuristic

[8]There is also a danger that each covariance on the RHS of (15) is nonzero but that by chance $\beta(\sigma_{wu}-\sigma_{wv})+(\sigma_{u\varepsilon}-\sigma_{v\varepsilon})=0$. If we adopt a Bayesian perspective and think of the covariances as free parameters drawn from some continuous distribution then this occurrence would be a set of measure zero. However and as is always the case with composite nulls like these moments may be such that the expression is close to zero and if so this will compromise small sample power.

[9]By examining potential differences in the β estimates delivered by each instrument in a set we are alluding to the use of a Wald test of instrument validity. Below we use the J (LM) test not the Wald test. However under the null, Wald and J are asymptotically equivalent. If Wald diverges (does not diverge) then J will also diverge (not diverge)

reasoning. Although here we test for instrument validity, we would still wish to propose instruments that by prior reasoning have the greatest chance of being valid. By doing so we minimise our exposure to type II error. With this in mind the instruments for x(for z) we consider in the first instance are a lagged value of z(of x) (where lags are available) and the level and third power of z(of x) respectively. Of course this list could be expanded (second lags, fifth powers etc) but the current choice is sufficient to highlight pitfalls and issues surrounding the use of such instruments in this context. The reason for making odd rather than even powers the first port of call in the search for relevant instruments is that the IV estimator will depend on the $n+1th$ moment of w where n is the power of the instrument. If this is zero - as would be the case if n is odd and w had a symmetric pdf - the estimator is not defined. We avoid using own lags as instruments; in many applications where lagged measures are available it is quite likely that the ME is autocorrelated which would invalidate such instruments a priori.[10] Finally in what follows we always include z in the list of instruments for x and vice versa as in practice it is very likely that they will be respectively strong instruments for each other. This inclusion makes uncorrelatedness of the ME's an intrinsic part of each null we test.

We now examine the limits of IV estimates of β derived from each of our suggested instruments. We only provide results for z being instrumented; the formulae and corresponding discussions may be symmetrically applied to the case where x is instrumented instead. Equations (16) to (21) below give the large sample estimates of β arising when we instrument z using a) x, b) x_{-1} and c) x^3, We label these B_{1x}, B_{2x} and B_{3x} respectively. The corresponding null restrictions that imply consistency, R_{1x}, R_{2x}, R_{3x} are given below each respective case. It will become clear that the lessons learnt from using x^3 and x_{-1} as instruments would also apply were we to expand the set of instruments to include higher

[10] Below we test for ME autocorrelation by regressing x-z on its lag. This test is similar in construction to T_C.

powers and higher order lags. In the formula below we use the fact that $\sigma_{w\varepsilon} = 0$.

$$B_{1x} = \beta + \frac{\sigma_{u\varepsilon} - \beta(\sigma_{vw} + \sigma_{uv})}{\sigma_{xz}} \tag{16}$$

$$R_{1x} : \sigma_{u\varepsilon} = \sigma_{uw} = \sigma_{uv} = 0 \tag{17}$$

$$B_{2x} = \beta + \frac{\sigma_{u_{-i}\varepsilon} + \sigma_{w_{-i}\varepsilon} - \beta(\sigma_{vw_{-i}} + \sigma_{u_{-1}v})}{\sigma_{x_{-i}z}} \tag{18}$$

$$R_{2x} : \sigma_{u_{-1}\varepsilon} = \sigma_{w_{-i}\varepsilon} = \sigma_{vw_{-i}} = \sigma_{u_{-i}v} = 0 \tag{19}$$

$$B_{3x} = \beta + \frac{\sigma_{w^3\varepsilon} + \sigma_{u^3\varepsilon} + 6\sigma_{wu^2\varepsilon} + 6\sigma_{w^2u\varepsilon} - \beta(\sigma_{u^3v} + \sigma_{w^3v} + 6\sigma_{u^2vw} + 6\sigma_{w^2uv})}{\sigma_{zx^3}} \tag{20}$$

$$R_{3x} : \sigma_{w^3\varepsilon} = \sigma_{u^3\varepsilon} = \sigma_{wu^2\varepsilon} = \sigma_{w^2u\varepsilon} = \sigma_{u^3v} = \sigma_{w^3v} = \sigma_{u^2vw} = \sigma_{w^2uv} = 0 \tag{21}$$

How reasonable a priori are the moment restrictions in equations (17), (19) and (21) and are any so unlikely to hold that we would not wish to subject them to a test in the first place? The moments may be split into two; those involving inter alia ε and w and those involving *only* u, v and w and their lags. In most cases a priori reasoning would not rule out the idea that u, v and w could be mutually uncorrelated and we might think that this reasoning should extend to independence. As for moments involving inter alia w and ε things are less favorable. Whilst we know that $\sigma_{w\varepsilon}$ is definitionally zero, $\sigma_{w^3\varepsilon}$ may not be so; if w was exogenous then we would need the relationship between y and w to be linear (or, in practice, approximately so) for it to hold. However it may be that whilst $\sigma_{w^3\varepsilon} = 0$ is violated mathematically, the violation is not severe enough to generate quantitatively important asymptotic estimation bias. Finally sufficient (but not necessary) conditions for $\sigma_{w_{-i}\varepsilon} = 0$ to hold are that the ME's are classical (an assumption separately examined by T_C) or that ε is non autocorrelated.
As we have seen already, if the only violation of the null was a nonzero correlation of u and v, then using x as an instrument for z and vice versa produces identically biased estimates. Therefore including the two corresponding orthogonality restrictions in the set to be tested will compromise test power. It is for this exact reason that we keep separate (T_x and T_z) rather than combine them into a single test. [11] To summarise whilst the restrictions above may have different degrees of prior credibility depending on the application, none are are so unreasonable that they should not even be subjected to a test.

Assuming that there are sufficient relevant instruments to compute either or both of

[11] In simulation results not reported here but available on request we use the setup in the next section to establish that a test sequence based on T_C and a second test that combines the restrictions in T_x and T_z offers considerably less power than the three test sequence proposed in this paper.

T_x and T_z and if they, together with T_C, fail to reject, we may then have enough confidence in the composite null underpinning the three tests (the union minus the intersection of their individual respective nulls) to combine the two sets of orthogonality conditions underpinning T_x and T_z to obtain a single efficient estimate of β via GMM.

Before moving on we note an important feature of our tests exposed by the formulae above. There are clearly some dimensions in which the null could fail that would not undermine consistency of β. For example were $\sigma_u^3 v = 0$ in R_3 to be the only element of the null to fail, the corresponding (asymptotic) rejection would lead to the false inference that consistent estimates of β are not obtainable when in fact they are. However, in most of the cases like this above, there are relatively few economic contexts where failure of the restriction would not go hand in hand with failure of uncorrelatedness. [12].Nonetheless we should still consider our test to be conservative; In some circumstances it will err on the side of finding against consistent estimation even when consistent estimation is possible.

5 A Simulation Study.

Above we proposed a diagnostic procedure based on the execution of three tests; T_C (14), T_x and T_z - assuming of course that suitable relevant instruments are available. If each test in the sequence fails to reject its null then combining the instruments involved in T_x and T_z in a single GMM procedure would produce an estimate of β that was credibly consistent. Here we assess the power of this diagnostic procedure to detect biases of various magnitudes arising when the composite null underlying the three tests is violated.

The test bed for our experiments is

$$y_i = \beta w_i + \varepsilon_i \tag{22}$$

$$w_i = \rho w_{i-1} + \xi_i \tag{23}$$

$$x_i = w_i + u_i \quad where \tag{24}$$

$$u_i = (u_i^* + \gamma_u \varepsilon_i + \delta_u \xi_i)/k_u \tag{25}$$

$$z_i = w_i + v_i \quad where \tag{26}$$

$$v_i = (v_i^* + \gamma_v \varepsilon_i + \delta_v \xi_i)/k_v \tag{27}$$

$$i = 1,,,N \tag{28}$$

[12] One key counterexample here may be ARCH behaviour in financial applications

Where $\varepsilon_i, u_i^*, v_i^*$ and ξ_i are iid mean zero exponential [13] with unit variance. The parameters k_x and k_z normalise the variance of u_i and v_i to one in each simulation. Asymptotically relevant instruments that we use to generate our tests statistics are z, z_{-1} and z^3 for x and z, $x_{,-1}$ and x^3 for z. [14] The compound null hypothesis of the test sequence is

$$\gamma_u = \gamma_v = \delta_u = \delta_v = 0 \tag{29}$$

Under this null, ME is classical and our instruments are valid. The most obvious problem in the context of a sequence of tests is controlling for size[15] of the procedure as a whole. If we are able to conduct the complete test sequence and if the tests were independent then using the standard significance levels of 5% would result in an overall size of the test procedure of about 14%. If the (three) tests were perfectly correlated the size would be 5%. This oversize problem is hardly unique to our current context; investgators have always used batteries of tests to assess the validity of their IV procedures. Nonetheless we propose to counter the the size issue by adopting confidence levels of 2.5% for each of the three tests.

Another issue - again not unique to our context - is that relevant instruments may not exist. In all of our simulations $x(z)$ is a strong instrument for $z(x)$. This is something we feel is likely to be a common occurrence in the data. The significance (relevance) of the lag and cube terms is a different matter. Although the experiments are configured so that these instruments are always strong in large samples, for small samples they may not be. Therefore in each experiment we compute an F test for the relevance of the additional instruments for x and z respectively. If one of these has a p-value above 5% [16] then we do not proceed to compute the corresponding J test and in that simulation the number of tests in the sequence is reduced from three to two. If neither F test for additional instruments passes at the 5% level then we do not proceed at all and that particular simulation will be dropped from the results. These actions mimic those that empirical investigators would take in any application - no investigator should invoke overidentification tests with weak

[13]Using exponential rather than normal pdf's and adopting the AR(1) structure for w was important to ensure that the relevance of these instruments.

[14]Including terms in higher powers and lags of x and z would preclude analysis of power for small N such as the case of N=50 we include below.

[15]Here "size" is the chance we reject the composite null of the three tests when that composite null is true.

[16]Switching to a criterion of 1% did not alter the power estimates much but it did reduce the proportion of cases we could execute our test sequence in small samples.

(additional) instruments. We re-emphasise that regardless of the existence of a sufficient number of relevant instruments T_C may always be executed. If this test fails to reject we may have some confidence in the hypothesis that ME's are classical in nature which in turn could reveal information about the likely sign and magnitude of the bias in β.

Before we discuss and defend our choice of $\gamma's$ and $\delta's$ used for the simulations we note an important feature of the current investigation. The null hypothesis is not of any intrinsic economic interest. By contrast the size of the estimation bias caused by its violation is of paramount concern. This in turn will depend on the estimation method used. Typically either Twostep GMM (TSGMM) or Iterated GMM are used in empirical work. Both methods offer asymptotically efficient estimates but for reasons of computational efficiency we use the former. For each parameter configuration under the alternative hypothesis therefore, we compute the (absolute value of the proportional) TSGMM asymptotic[17] bias. This helps us assess the criticality of power in each parameter scenario; there is little cost of low power in scenarios where estimation bias is also low but high cost in scenarios where bias is high.

Because of the normalisations in (25) and (26) the signal to noise ratio in x and z is constant. The standard deviation of ME is about 40% of that of w. This roughly matches the ME reported for median size states in the annual LAUS data we use in the empirical application below (see https://www.bls.gov/lau/lastderr.htm). In sensitivity analyses we found that moderate increases/decreases in the ME standard deviation had little effect on the results we present here. We present reults for eleven parameter configurations and five sample sizes. The parameter configurations are represented by the row vector $P(\gamma_u, \gamma_v, \delta_u, \delta_v)$. There are three qualitatively distinct cases.; a) the cases where violations of the null are due entirely to common components in the two ME's - P_2, P_6, P_7 and P_{11} in the table below, b) the cases where violations of the null are only partially due to common components in the two ME's - P_4, P_5, P_9 and P_{10} and c) where violations of the null are due to components that are not common to the two ME's - P_3 and P_8. Finally scenario P_1 estimates size of the test sequence. The results for the 11 configurations are presented in Table 1.

The results for P_1 show that test size in large samples is marginally oversized at around 7% but my be as high as 10% in small samples. Turning to power, cases in a) - although rather pathological - allow us to assess what happens when T_C is inconsistent

[17] An alternative is to estimate the average bias for each sample size. As a sensitivity check we did this for $N = 50$ in cases P_6 and P_{11}. The results were little different to the asymptotic biases given in the Table.

and the overall power of the sequence rests largely on T_X and T_Z. We see that power is quite low for sample sizes below 100. Whilst this is of only modest concern in P_2 and P_6 where asymptotic bias is relatively low, it is of serious concern in P_7 and P_{11}. In the latter two cases the J tests (T_X and T_Z) seem to require sample sizes of around 500 to have reasonable power. However cases b) and c) display reasonable power even in small sample sizes.

Table 1: Simulation Results

		N				
		50	100	250	500	1000
$P_1 = (0, 0, 0, 0)$	Bias = 0%	9.3	7.9	7.2	7.2	7.0
$P_2 = (.5, .5, .5, .5)$	Bias = 12%	18.4	21.6	33.0	57.4	88.0
$P_3 = (.5, -.5, .5, -.5)$	Bias = 2%	99.1	100	100	100	100
$P_4 = (.5, -.5, .5, .5)$	Bias = 23%	75.2	100	100	100	100
$P_5 = (.5, .5, .5, -.5)$	Bias = 24%	57.2	87.4	100	100	100
$P_6 = (-.5, -.5, -.5, -.5)$	Bias = 18%	8.8	17.5	22.5	53.3	79.4
$P_7 = (1, 1, 1, 1)$	Bias = 19%	33.2	45.6	79.2	95.3	100
$P_8 = (1, -1, 1, -1)$	Bias = 2%	100	100	100	100	100
$P_9 = (1, -1, 1, 1)$	Bias = 44%	97.2	100	100	100	100
$P_{10} = (1, 1, 1, -1)$	Bias = 48%	100	100	100	100	100
$P_{11} = (-1, -1, -1, -1)$	Bias = 39%	17.1	32.7	67.7	93.5	100

Notes: "Bias" is the absolute asymptotic TSGMM bias of each of the 11 Cases. N is the number of observations and pr is the proportion of times at least one of T_C, T_x and T_z reject at the 2.5% level of significance.

It is hard to be definitive from any numerical exercise. But these simulations indicate that - apart from pathological cases where violations of the null are entirely down to common ME components - the test sequence we propose is a useful diagnostic tool even in samples as small as 50.

6 Two Empirical Applications

We apply our test sequence and proposed estimations to two empirical scenarios. First we analyse the two wage cyclicality estimates obtained from using HP filtered GDP and the aggregate unemployment rate and second we analyse the comovement of burglaries with state level unemployment rates in the US as measured by the LAUS. The latter exercise highlights the ME issues arising when using state unemployment measures from the the LAUS. There we artificially generate two noisy measures of state unemployment by randomly assigning county level unemployment rates to two subsamples. The former application is more of a latent variable exercise. It asks the question "given that the business cycle is an intrinsically unobserved variate, can we reconcile two estimates of wage cyclicality from two distinct measures of the cycle?"

6.1 The Cyclical Component of Incumbent Wages

There is a huge literature on estimating the cyclical component of wages (See for example Barsky et al. (1994) and, more recently, Devereux and Hart (2006)). In most of these studies the aggregate rate of unemployment is used as the cyclical measure. However the business cycle is an intrinsically unobserved variable. For example GDP may be used to derive a cyclical indicator rather than unemployment. Here we revisit this literature by treating HP filtered GDP and aggregate unemployment as two separate noisy measures of the business cycle. Strictly speaking this is more in the spirit of a latent variable application than one of ME but we adapt it to our template.

We treat the aggregate unemployment rate and HP filtered log GDP as noisy measures of the cycle. Following the literature, we estimate a specification in first differences and refer to the change in unemployment as x, the change in the HP variate as $-z$, and the composition corrected aggregate real wage growth as y.[18] We normalise z to have the same standard deviation as x and take its negative value. This ensures that the comovement of wage growth with x is comparable with that of z and that β is scaled in the traditional way as a cycical response of real wages to percentage point changes in the unemployment rate.

We wish to obtain a composition-bias-free estimate of aggregate real wage growth. To this end we use the hourly real wage data of stayers from the PSID between 1980 and 2019 to estimate a Mincer regression[19] of wage growth on a quadratic in tenure and

[18] We use a smoothing parameter of 6.25 which Ravn and Ulhig (2002) argue is optimal for annual US GDP data. We refer in the text to "real wage growth". More formally this is the change in the lof of real wages. We deflate wages by the CPI.

[19] We use the PSID weights in this first stage and focus solely on the wages of heads of households. We drop cany hange in log wage that exceeds 1.0 from the sample. In 1998 the survey becomes bienial. For those later years we take changes over two years divided by two for all of the regressions. This weakens

experience (measured by worker age minus 16) plus year and state fixed effects. The year fixed effects are then extracted and in a second stage used as the y variable. This variable is a time series of arguably composition-free aggregate wage growth. In a second stage we use this time series to estimate cyclicality via OLS, IV and GMM - as per the analysis above - using the number of data points in each yearly cross section as weights. (See Stockman (1983) and Barsky et al. (1994), who also use this two step method). [20] We apply the HP filter to the log of real GDP and then first difference to obtain z. Finally we should note that the PSID wage measure (and many other survey based wage measures) is itself subject to a high degree of ME. Referring to validation studies, Kim and Solon (2005) argue that the ME in the PSID reverts to mean. They suggest that PSID derived cyclicality estimates have to be inflated by 50% to compensate. The implications for our application are not severe; ME of this nature impinges equally on estimates of β from both of our cycle measures and we may adjust the final single GMM estimate of β we derive appropriately to allow for it.

A major issue with this literature is the scarcity of annual time series observations available - the 30 or so we have here is actually large relative to many studies in the area. One consequence is that test power may be low so results here should be interpreted with caution. More directly it means we cannot undertake a full exploration of relevant instruments in the manner of the last section. Instead we estimate two first stage regressions each for $x(z)$ - one using z and z_{-1} (x and x_{-1}) as instruments and the other using z and z^3 (x and x^3) as instruments respectively. As before we keep the level of $x(z)$ but select the extra instrument (lag or third power) on the basis of statistical significance. The first four lines of Table 3 below give the relevant t-tests denoted intuitively as t_{ab} where a is the variable being instrumented and b is the instrument. The results show clearly that for x the lag is preferred over the cube whilst for z it is the cube term that dominates. The results for the corresponding J tests and for T_C are given in lines 5 to 7 of Table 2.

the strength of lagged instruments somewhat but we see below that a lagged HP term is a highly relevant instrument for unemployment.

[20] If the number of annual cross sectional data points is large then this method delivers the same estimate of wage cyclicality as direct panel estimation. It is computationally convenient here where we use a GMM estimator.

Table 2: The Cyclicality of Real Wages

	Test Value	p-Value
$t_{xz_{-1}}$	2.621	0.015
t_{xz^3}	0.902	0.376
$t_{zx_{-1}}$	1.748	0.094
t_{zx^3}	2.386	0.025
T_C	0.877	0.387
T_ρ	0.177	0.865
T_x	0.876	0.349
T_z	0.069	0.793
	Estimates of β	
OLS(x)		−0.728
		(0.257)
OLS(z)		−0.551
		(0.242)
GMM		−0.891
		(0.204)

Notes: Robust standard errors in parentheses.

Each of the overidentification tests passes very comfortably. Even factoring in the small sample size these test statistics have very high p-values. This offers some reassurance that IV delivers consistent estimation of β. The bottom half of the table gives the estimates of wage cyclicality. The results from OLS conflict; whilst both are significant the standard (unemployment rate) semi elasticity is .72 whilst the estimate from the HP cyclicality measure is around .55. The GMM estimate is larger (in absolute value) than both the OLS estimates - substantially so in the case of the HP measure.

6.2 The Comovement of Within State Unemployment and Burglaries

There has always been an interest in the links between economically motivated crime and unemployment. A notable paper in this area is Raphael and Winter-Ebmer (2001). Using data from 1971 to 1997 and a specification in levels, they find a significant -arguably causal - link between property crime and state level unemployment using state level oil shocks as instruments for unemployment. Here we are less ambitious. We merely want to use this context as a test bed for the impact of measurement error on the comovement of state unemployment and state crime and will not attempt to identify any causal effects. However, and in contrast to Raphael and Ebmer, we adopt a first differences specification to remove state fixed effects and stochastic and determininstic trends whose presence may otherwise lead to spurious correlation[21].

The FBI collect three measures of economically motivated crime; Auto Theft, Robberies and Burglaries. We focus on the last of these because burglaries are arguably more driven by local economic factors. By contrast Auto Theft and Robberies are increasingly dominated by organised crime networks (see for example the discussion in Longman (2006) and may therefore be more related to US wide factors[22]. As we have noted already the LAUS are subject to severe ME with the standard deviation of ME often being as much as 40% of that of the true measure. This and the biases it implies are rarely acknowledged in the literature. [23] The exercise here is to assess its impact on the OLS estimate of state unemployment on burglaries.

To obtain two measures of state unemployment we split each state's counties into two bins according to their position in the alphabet making sure that each bin ends up with approximately the same population.[24] The unemployment rate for each of the two bins in each year is computed, first differenced and - to keep the core analysis of the paper firmly to the fore - we call these measures x and z. The (change in log of the) number of burglaries we denote by y and (the change in the) true unobserved state unemployment rate by w. It is unlikely that the position in the alphabet of county names is related to either burglaries or economic outcomes so our splitting of counties into two groups is

[21] Adding year fixed effects to a levels specification may control for a single stochastic trend but it will not be effective if two or more such trends are present.

[22] Experimentation with Auto Theft and Robberies suggested no relationship with state level unemployment rates whatsoever; coefficients varied in sign, were closer to zero and less significant than the OLS results for Burglaries. Of course if β really was zero as may be the case for these two crimes, there is no attenuation bias to worry about.

[23] For an example of a recent use of LAUS see Grigsby et al. (2021) Using LAUS' state unemployment measure as a regressor they find very low wage-unemployment elasticities and little difference between new hires and incumbents. This may in part be due to attenuation bias arising because of the high degree of ME in the LAUS statistics.

[24] For the first bin - leading to measure x, we start with the county earliest in the alphabet and keep adding counties in alphabetical order until the population covered reaches (or exceeds) one half the state total for that year. The remining counties go into the second bin - leading to measure z.

likely to be "at random"[25]

If we observed w we would aim to estimate

$$y_{it} = \alpha + \beta w_{it} + YFX + \varepsilon_{it}$$
$$i = 1, ..50..t = 1991...2019$$

for states i[26] and years t by OLS where "YFX" denotes year fixed effects, included to absorb US wide macro effects particularly the aggregate rate of unemployment. The set of instruments used follows the procedure used in the simulations. We regress $x(z)$ on the level, the lag and cube of $z(x)$ The level of $x(z)$ is always included but We only include the additional two instruments if they are jointly significant at the 5% level. The two F tests for the latter, F^x and F^z , are given in the first two lines of Table 3. [27] The results show that whilst the extra two instruments are highly significant in explaining z there seem to be no overidentifying instruments available for z. Whilst it is hard to explain why the two series display such different properties it is not fatal for our procedure; we may still evaluate T_C and T_x. The third and fifth lines of Table 3 respectively give these tests. For completeness we also test for first order autocorrelation of the ME's. [28] To do so we follow the logic behind T_C and regress $x - z$ on its lag. T_ρ in line four of the table gives the corresponding t-ratio and p-level. The remaining lines in the table give the estimates of β obtained using OLS on the measures and the GMM estimate obtained combining the two sets of orthogonality conditions.

[25] An alternative would have been to use computer generated random numbers to split counties. However the results from this procedure would depend on our own choice of random number seed. Using the alphabet is more transparent and facilitates replication.

[26] The state index i is incomplete; it excludes Alaska, District of Columbia and Rhode Island because these states had insufficient counties to conduct a randomised split.

[27] Year fixed effects are removed from x, and z prior to transformations.

[28] If the ME's are white noise then own lagged values are also potentially valid instruments

Table 3: Burglaries and State Unemployment Rates

	Test Value	p-Value
$F^x_{2,27}$	0.460	0.637
$F^z_{2,27}$	5.900	0.0075
T_C	1.022	0.316
T_ρ	1.252	0.221
T_x	3.293	0.193
Estimates of β		
OLS(x)		0.502
		(0.424)
OLS(z)		0.364
		(0.437)
GMM		1.072
		(0.412)

Notes: Year fixed effects included. Standard errors are clustered by state.[a]

[a]In the Annex we discuss the use of clustered standard errors in our current context. The assumptions required for clustered standard errors to offer valid inference are stronger than those normally required.

The test for Classical ME (T_C) passes comfortably with a p value around .6. The autocorrelation test has a p-value of around .2 which is quite low. This justifies to some extent our use of x_{-1} (z_{-1}) as an instrument for $z(x)$ rather than $z_{-1}(x_{-1})$ - autocorrelation of ME's would invalidate the use of own lags as instruments a priori. The J-test of instrument validity (T_x, here a χ^2_2) passes comfortably. The OLS results display coefficients of around .5 but are not particularly significant. The estimate of β from GMM is 1.07 and significant. The low p-value for T_C combined with the relatively low and insignificant (high and significant) OLS(GMM) estimate are strongly indicative of classical attenuation bias. In sum, these results indicate significant comovement of (the change in the log of) burglaries with the (change in the) state unemployment rate - a comovement not picked up in OLS regressions[29]. These results and those for wage cyclicality above all

[29]The OLS estimate using the LAUS measure itself was - like those for x and z - around .5 and insignificant. We expand on this point below in the summary and conclusion

point in the same direction. There seems to be attenuation bias in OLS estimates of β in both applications. Tests indicate validity of instruments (x for z and/or vice versa) and use of these instruments raises the estimated β substantially.

7 Summary and Conclusion

In this paper we have proposed a sequence of tests and estimations to gain consistent estimates of the OLS parameter (β) of y on an unobserved regressor w using two noisy measures of the regressor x and z. The approach is novel because the proposed GMM procedure uses functions of x as instruments for *z and simultaneously vice versa*. The three test statistics have a composite null (the union minus the intersection of the three respective nulls of instrument validity) whose "truth" ensures that system GMM consistently estimates β. In a standard setting - one with heterogeneous responses - overidentification tests cannot be used to assess instrument validity because different instruments lead to different weighted averages of the responses even where the instruments themselves are valid. Here our instruments are targeting a single entity, namely β the OLS estimate of y on w so this issue does not arise. Indicative simulations suggest the test sequence has reasonabbe power even in small samples except in the somewhat pathological case where the violation of the null is due solely to the existence of common additive components in the two ME's. Two empirical applications - one estimating the co-movement of burglaries and state level unemployment and the other estimating the co-movement of wages with a business cycle latent variable - illustrate the procedure. In the applications we find strong evidence of attenuation bias. The final estimates are higher in absolute value in both applications suggesting that our procedure corrects for attentuation bias or at least attenuates it.

We close with a note about the random sampling of county unemployment rates in the burglary application. In results not presented here, we found that OLS estimates using the standard state LAUS measures are similar in size and significance to those from the two constructed measures. By contrast our use of IV with these two measures raises the estimated β. This suggests that when using LAUS data as regressors one could potentially improve the estimated regression coefficients with some kind of random sampling such as that attempted here. Of course the sampling method we chose - based on alphabetical ordering of counties - is not unique. It would be interesting to analyse the use of a more systematic method based on a complete set of random samples from the counties in a jacknife-style procedure. Future work may go in this direction.

8 Annex

8.1 Allowing for Controls

To conserve notation we amend the model to allow for a single control variable q_i but the arguments easily extend to multiple (linear) controls). We wish to orthogonalise the instruments and y with respect to q. The simplest way to achieve this would be to regress each variable on q and use the residuals for the analysis. However we are using powers of x and z as instruments and it may be preferable to remove the first order effect of q from x and z before transforming them - as we do in the empirical application. This requires us to make stronger assumptions than would a standard orthogonalisation. The assumptions are given below. Intuitively they require that q enters the model only via a linear term. [30]Having said all of this, the results below are easily adapted to the simpler case of orthogonalising all the instruments used directly.

Equations (54) to (57) give the amended model plus auxiliary assumptions..The idea is to show that if we replace x, z etc with their orthogonalised counterparts, our original orthogonality conditions have exactly the same form. Equation (57)((58)) defines $w_i^{\perp}(y_i^{\perp})$ as the large sample orthogonal projection of $w(y)$ on q. As noted above the auxiliary assumptions in (39) effectively say that only q itself enters the model. Equations (59) to (61) define the finite sample orthogonal projections of y, x and z respectively on w in terms of their large sample counterparts plus the sampling errors. Our intention is to use $\widehat{x}$ and $\widehat{z}$ (and their lags and cubes) together with $\widehat{y}$ in exactly the same way as we used x, z and y in the text. That is we show that in large samples the results derived in the text for x, z and y also hold true for $x_i^{\perp}, z_i^{\perp}$ and $y_i^{\perp}$.

[30] The results here generalise to many controls so adding lags or higher order terms in q - where deemed necessary - is not an issue.

$$y_i = \beta w_i + \gamma q_i + \varepsilon_i \tag{30}$$

$$x_i = w_i + u_i \tag{31}$$

$$z_i = w_i + v_i \tag{32}$$

$$w_i = \alpha q_i + w_i^{\perp} \tag{33}$$

$$y_i^{\perp}\{ = y_i - \delta q_i\} = \beta w_i^{\perp} + \varepsilon_i \quad where \quad \delta = \alpha\beta \tag{34}$$

$$\widehat{y_i} = y_i - \widehat{\delta} q_i = y_i^{\perp} + (\delta - \widehat{\delta}) q_i \tag{35}$$

$$\widehat{x}_i = x_i - \widehat{\alpha_x} q_i = x_i^{\perp} + (\alpha - \widehat{\alpha_x}) q_i = w_i^{\perp} + (\alpha - \widehat{\alpha_x}) q_i + u_i \tag{36}$$

$$\widehat{z}_i = z_i - \widehat{\alpha_z} q_i = z_i^{\perp} + (\alpha - \widehat{\alpha_z}) q_i = w_i^{\perp} + (\alpha - \widehat{\alpha z}) q_i + v_i \tag{37}$$

$$where \quad plim(\widehat{\alpha_x}) = plim(\widehat{\alpha_z}) = \alpha \quad and \quad plim(\widehat{\delta}) = \delta \tag{38}$$

$$Assumption^{31} : u_i, v_i \ \ and \ \ w_i^{\perp} \ \ are \ independent \ of \ q \tag{39}$$

$$N^{-\frac{1}{2}} \widehat{x}_i (\widehat{y}_i - \beta \widehat{z}_i) \tag{40}$$

$$= N^{-\frac{1}{2}} x_i^{\perp} (y_i^{\perp} - \beta z_i^{\perp}) + A \quad where \tag{41}$$

$$A = \{N^{\frac{1}{2}}(\alpha - \widehat{\alpha_x}) N^{\frac{1}{2}}(\delta - \widehat{\delta})\}\{N^{-\frac{3}{2}} q_i^2\} + \{N^{\frac{1}{2}}(\alpha - \widehat{\alpha_x})\}\{N^{-1} y_i^{\perp} q_i\} \tag{42}$$

$$-\beta\{N^{\frac{1}{2}}(\alpha - \widehat{\alpha_x}) N^{\frac{1}{2}}(\alpha - \widehat{\alpha_z})\}\{N^{-\frac{3}{2}} q_i^2\} - \beta\{N^{\frac{1}{2}}(\alpha - \widehat{\alpha_x})\}\{N^{-1} z_i^{\perp} q_i\} \tag{43}$$

$$\tag{44}$$

$$N^{-\frac{1}{2}} \widehat{x}_{i-1} (\widehat{y}_i - \beta \widehat{z}_i) \tag{45}$$

$$= N^{-\frac{1}{2}} x_{i-1}^{\perp} (y_i^{\perp} - \beta z_i^{\perp}) + B \quad where \tag{46}$$

$$B = \{N^{\frac{1}{2}}(\alpha - \widehat{\alpha_x}) N^{\frac{1}{2}}(\delta - \widehat{\delta})\}\{N^{-\frac{3}{2}} q_i q_{i-1}\} + \{N^{\frac{1}{2}}(\alpha - \widehat{\alpha_x})\}\{N^{-1} y_i^{\perp} q_{i-1}\} \tag{47}$$

$$-\beta\{N^{\frac{1}{2}}(\alpha - \widehat{\alpha_x}) N^{\frac{1}{2}}(\alpha - \widehat{\alpha_z})\}\{N^{-\frac{3}{2}} q_i q_{i-1}\} - \beta\{N^{\frac{1}{2}}(\alpha - \widehat{\alpha_x})\}\{N^{-1} z_i^{\perp} q_{i-1}\} \tag{48}$$

$$N^{-\frac{1}{2}} \widehat{x}_i^3 (\widehat{y}_i - \beta \widehat{z}_i) = \tag{49}$$

$$N^{-\frac{1}{2}} x_i^{\perp 3} (y_i^{\perp} - \beta z_i^{\perp}) + C \quad where \tag{50}$$

$$C = \{[N^{\frac{1}{2}}(\alpha - \widehat{\alpha_x})]^3 N^{-2} q_i^3 + 6[N^{\frac{1}{2}}(\alpha - \widehat{\alpha_x})]^2 N^{-\frac{3}{2}} x_i^{\perp} q_i^2 + \mathbf{6[N^{\frac{1}{2}} (\alpha - \widehat{\alpha_x})] N^{-1} x_i^{\perp 2} q_i\}} \tag{51}$$

$$\{\boldsymbol{y}_i^{\perp} + N^{\frac{1}{2}}(\delta - \widehat{\delta}) N^{-\frac{1}{2}} q_i - \boldsymbol{\beta z}_i^{\perp} - \beta N^{\frac{1}{2}}(\alpha - \widehat{\alpha_z}) N^{-\frac{1}{2}} q_i\} + \tag{52}$$

$$\boldsymbol{x}_i^{\perp 3} \mathbf{\{N^{\frac{1}{2}} (\delta - \widehat{\delta}) N^{-1} q_i - \beta N^{\frac{1}{2}} (\alpha - \widehat{\alpha_z}) N^{-1} q_i\}} \tag{53}$$

[31] Independence is stonger than is needed here, However we do require a slew of zero correlations between lags and powers of u, v, w and q so adopting the independence assumption adds clarity..

$$
\begin{aligned}
y_i &= \beta w_i + \gamma q_i + \varepsilon_i && (54)\\
x_i &= w_i + u_i && (55)\\
z_i &= w_i + v_i && (56)\\
w_i &= \alpha q_i + w_i^{\perp} && (57)\\
y_i^{\perp} &= y_i - \delta q_i = \beta w_i^{\perp} + \varepsilon_i, \qquad \text{where } \delta = \alpha\beta && (58)\\
\widehat{y}_i &= y_i - \widehat{\delta} q_i = y_i^{\perp} + (\delta - \widehat{\delta}) q_i && (59)\\
\widehat{x}_i &= x_i - \widehat{\alpha_x} q_i = w_i^{\perp} + (\alpha - \widehat{\alpha_x}) q_i + u_i && (60)\\
\widehat{z}_i &= z_i - \widehat{\alpha_z} q_i = w_i^{\perp} + (\alpha - \widehat{\alpha_z}) q_i + v_i && (61)\\
\text{where} \quad & \operatorname{plim}(\widehat{\alpha_x}) = \operatorname{plim}(\widehat{\alpha_z}) = \alpha, \quad \operatorname{plim}(\widehat{\delta}) = \delta && (62)
\end{aligned}
$$

Assumption [32]: u_i, v_i, and $w_i^{\perp}$ are independent of q

$$
\begin{aligned}
N^{-1/2} \widehat{x}_i (\widehat{y}_i - \beta \widehat{z}_i) &= N^{-1/2} x_i^{\perp} (y_i^{\perp} - \beta z_i^{\perp}) + A, \qquad \text{where} && (63)\\
A &= N^{1/2} (\alpha - \widehat{\alpha_x}) N^{1/2} (\delta - \widehat{\delta}) N^{-3/2} q_i^2 && (64)\\
&+ N^{1/2} (\alpha - \widehat{\alpha_x}) N^{-1} \sum y_i^{\perp} q_i && (65)\\
&- \beta N^{1/2} (\alpha - \widehat{\alpha_x}) N^{1/2} (\alpha - \widehat{\alpha_z}) N^{-3/2} \sum q_i^2 && (66)\\
&- \beta N^{1/2} (\alpha - \widehat{\alpha_x}) N^{-1} \sum z_i^{\perp} q_i && (67)
\end{aligned}
$$

$$
\begin{aligned}
N^{-1/2} \widehat{x}_{i-1} (\widehat{y}_i - \beta \widehat{z}_i) &= N^{-1/2} x_{i-1}^{\perp} (y_i^{\perp} - \beta z_i^{\perp}) + B, \qquad \text{where} && (68)\\
B &= N^{1/2} (\alpha - \widehat{\alpha_x}) N^{1/2} (\delta - \widehat{\delta}) N^{-3/2} \sum q_i q_{i-1} && (69)\\
&+ N^{1/2} (\alpha - \widehat{\alpha_x}) N^{-1} \sum y_i^{\perp} q_{i-1} && (70)\\
&- \beta N^{1/2} (\alpha - \widehat{\alpha_x}) N^{1/2} (\alpha - \widehat{\alpha_z}) N^{-3/2} \sum q_i q_{i-1} && (71)\\
&- \beta N^{1/2} (\alpha - \widehat{\alpha_x}) N^{-1} \sum z_i^{\perp} q_{i-1} && (72)
\end{aligned}
$$

$$
N^{-1/2} \sum \widehat{x}_i^3 (\widehat{y}_i - \beta \widehat{z}_i) = N^{-1/2} \sum x_i^{\perp 3} (y_i^{\perp} - \beta z_i^{\perp}) + C, \qquad \text{where} \qquad (73)
$$

$$
\begin{aligned}
C &= \left[N^{1/2} (\alpha - \widehat{\alpha_x}) \right]^3 N^{-2} q_i^3 && (74)\\
&+ 6 \left[N^{1/2} (\alpha - \widehat{\alpha_x}) \right]^2 N^{-3/2} x_i^{\perp} q_i^2 && (75)\\
&+ 6 N^{1/2} (\alpha - \widehat{\alpha_x}) N^{-1} x_i^{\perp 2} q_i && (76)\\
&\times \left[y_i^{\perp} + N^{1/2} (\delta - \widehat{\delta}) N^{-1/2} q_i - \beta z_i^{\perp} - \beta N^{1/2} (\alpha - \widehat{\alpha_z}) N^{-1/2} q_i \right] && (77)\\
&+ x_i^{\perp 3} \left[N^{1/2} (\delta - \widehat{\delta}) N^{-1} q_i - \beta N^{1/2} (\alpha - \widehat{\alpha_z}) N^{-1} q_i \right] && (78)
\end{aligned}
$$

Equations (40),(45) and (49) are the counterparts to the empirical moments in the text where x, x_{-1} and x^3 are used to instrument for z. Symmetric analogue forms that arise when we use z to instrument x follow straightforwardly. Equations (63),(68) and (73) recast the moments in the text in terms of the large sample orthogonalised variates plus A, B and C respectively but it is easy to see that the latter disappear asymptotically. Under assumption (39) we can see clearly that A and B are $o(1)$. As for C inspection of each product shows that the highest order of probability terms arise from the summed products of the terms in bold in (51) and (52) and all the terms in (78). The assumption (39) ensures that these terms too are $o(1)$.

8.2 Nonsingularity of the Variance of the Orthogonality Conditions

GMM requires nonsingularity of the covariance matrix of the orthogonality conditions which we call V. With a slight abuse of notation in this subsection we denote the $Nx1$ vector of observations on x_i as x etc. Then the matrix V can be written as

$$\begin{aligned}
V &= E\{QS^2Q'\} \quad where \\
Q' &= [x'|x'_{-1}|x^{3'}|z'|z'_{-1}|z^{3'}] \\
S &= \begin{matrix} A & 0 \\ 0 & B \end{matrix} \\
& \quad where\ A\ and\ B\ are\ the\ following\ 3Nx3N\ matrices \\
A &= diag(\varepsilon - \beta u)\ and\ B = diag(\varepsilon - \beta u)
\end{aligned}$$

Unfortunately computing V is not straightforward in our framework. Even if the elements in $\varepsilon - \beta u$ are iid we cannot write $V = E\{Q[E\{S^2\}]Q'\}$. However we can check for nonsingularity in the simple but canonical case when $Q' = [x'|z']$. If V is nonsingular here we may be confident that any failure of the assumption to hold in our more general settings is unlikely to be a consequence of our use of reflexive instruments. In this case V is

$$V = Vcov\{N^{-\frac{1}{2}}x_i(y_i - \beta z_i)|N^{-\frac{1}{2}}z_i(y_i - \beta x_i)\}'$$

To illustrate, we show that V is nonsingular under the (strong) assumption that u, v

and ε are *iid*. In this case

$$
\begin{aligned}
V &= \begin{pmatrix} A+B & A \\ A & A+C \end{pmatrix} \\
\textit{where } A &= \beta^2\sigma_{vv}\sigma_{uu} + \sigma_{ww}\sigma_{\varepsilon\varepsilon} \\
B &= \beta^2\sigma_{uu}\sigma_{ww} + \sigma_{vv}\sigma_{\varepsilon\varepsilon} \\
C &= \beta^2\sigma_{vv}\sigma_{ww} + \sigma_{uu}\sigma_{\varepsilon\varepsilon}
\end{aligned}
$$

As both B and C are both strictly positive V here is positive definite. Obviously the *iid* assumption is very strong but we use it to make a point; If in any application V does turn out to be nonsingular it would not be a generic consequence of the use of reflexive instruments.

8.3 Clustered Standard Errors

To be able to estimate V (and hence compute standard errors) we need to impose a sufficient number of zeros on the cross correlations of elements in the orthogonality conditions. The industry standard method is to impose these zeros via a clustering assumption on the equations' errors. An implication of standard clustering is that if a_k and b_l are respective elements in the two othogonality conditions $N^{\frac{1}{2}}\sum_{i=1}^{N} a_i$ and $N^{\frac{1}{2}}\sum_{i=1}^{N} b_i$ that sit in different clustering units (e.g. different firms k and l or time periods k and l) then $E\{a_k b_l\} = 0$ is being assumed/imposed. Clustering also imposes the same condition on a_k and a_l i.e. on elements *within* an orthogonality condition. For our choice of instruments it is easy to show that a sufficient condition for clustering to be valid is that u, v and ε are independent of each other and of w *across* different cluster units. Whilst this condition is slightly stronger than is needed, distinguishing between it and the minimal level of dependence we need to hold in any empirical application is unlikely to be a desirable exercise.

References

Anderssen and Moen (2015). A simple improvement of the IV estimator for the classical errors-in-variables problem. *Oxford Bulletin of Economics and Statistics*, 78(1):126–144.

Angrist, J. and Imbens, G. (2000). The interpretation of instrumental variables estimators in simultaneous equations models wiht an application to the demand for fish. *Review of Economic Studies*, 67(3):499–527.

Barsky, R., Parker, J. A., and Solon, G. (1994). Measuring the cyclicality of real wages: How important is composition bias? *The Quarterly Journal of Economics*, 109(1):1–25.

Bound, J., Brown, C., and Mathiowetz, N. (2001). Measurement error in survey data. *Handbook of Econometrics*, 5:3705–3843.

Chen, X., Hong, H., and Nekipelov, D. (2011). Nonlinear models of measurement errors. *Journal of Economic Literature*, 49(4):901–937.

Dagenais, M. G. and Dagenais, D. L. (1997). Higher moment estimators for linear regression models with errors in variables. *Journal of Economietrics*, 76(1-2):193–221.

Devereux, P. J. and Hart, R. A. (2006). Real wage cyclicality of job stayers, within-company job movers, and between-company job movers. *Industrial Labor Relations Review*, 60(1):105–119.

Erickson and Whited (2002). Two-step GMM estimation of the errors-in-variables model using higher order moments. *Econometric Theory*, 18(3).

Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society*, 222:594–604.

Grigsby, J., Hurst, E., and Yildirmaz, A. (2021). Aggregate nominal wage adjustments: New evidence from administrative payroll data. *American Economic Review*, 111(2):428–471.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, 50(4):1029–1054.

Heckman, J. (1999). Instrumental variables: Response to Angrist and Imbens. *Journal of Human Resources*, 34(4):828–837.

Imbens, G. and Rubens, M. (2015). Causal inference for staistics, social and biomedical sciences: An introduction. *Cambridge University Press.*

Kim, B. and Solon, G. (2005). Implications of mean-reverting measurement error for longitudinal studies of wages and employment. *Review of Economics and Statistics*, 87(1):193–196.

Lewbel, A. (1997). Constructing instruments for regressions with measurement error when no additional data are available, with an application to patents and r&d. *Econometrica*, 65(5):1201–1213.

Longman, M. (2006). *Chapter 1 - The Problem of Auto Theft*, pages 1–21. Academic Press, Cambridge, MA.

Pesaran, M. H. (2015). Testing weak cross-sectional dependence in large panels. *Econometric Reviews*, 34(6-10):1089–1117.

Raphael, S. and Winter-Ebmer, R. (2001). Identifying the effect of unemployment on crime. *The Journal of Law Economics*, 44(1):259–283.

Ravn, M. O. and Ulhig, H. (2002). On adjusting the hodrick–prescott filter for the frequency of observations. *Review of Economics and Statistics*, 84(2):371–376.

Stockman, A. (1983). Aggregation bias and the cyclical behaviour of real wage. Unpublished.

Trapani, L. (2016). Testing for (in)finite moments. *Journal of Econometrics*, 191:57–68.

Wilhelm, D. (2019). Testing for the presence of measurement error. *CEMMMAP discusison paper*, CMP48/19.