Unbalanced Regressions and the Predictive Equation*

Daniela Osterrieder ‡§ Daniel Ventosa-Santaulària¶ J. Eduardo Vera-Valdés §

January 29, 2015

Abstract: Predictive return regressions with persistent regressors are typically plagued by (asymptotically) biased/inconsistent estimates of the slope, non-standard or potentially even spurious statistical inference, and regression unbalancedness. We alleviate the problem of unbalancedness in the theoretical predictive equation by suggesting a data generating process, where returns are generated as linear functions of a lagged latent $I(0)$ risk process. The observed predictor is a function of this latent $I(0)$ process, but it is corrupted by a fractionally integrated noise. Such a process may arise due to aggregation or unexpected level shifts. In this setup, the practitioner estimates a misspecified, unbalanced, and endogenous predictive regression. We show that the OLS estimate of this regression is inconsistent, but standard inference is possible. To obtain a consistent slope estimate, we then suggest an instrumental variable approach and discuss issues of validity and relevance. Applying the procedure to the prediction of daily returns on the S&P 500, our empirical analysis confirms return predictability and a positive risk-return trade-off.

Keywords: Predictive regression, persistent predictor, fractional integration, regression unbalancedness, IV estimation

JEL codes: G17, C22, C26, C58

---

*The authors acknowledge support from CREATES - Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation.

‡*Corresponding Author*: Rutgers Business School - Newark & New Brunswick, Department of Finance and Economics, Rutgers - The State University of New Jersey, 1 Washington Park, Newark, NJ 07102.
Email: dosterrieder@business.rutgers.edu
Phone: +1 973 353 2728

§CREATES and Aarhus University
¶Centro de Investigación y Docencia Económicas (CIDE)

1 Introduction

Returns on financial markets are risky. Investors in financial markets are uncertain about the future value of their investment. Modern portfolio theory (Markowitz, 1952) and the Capital Asset Pricing Model (CAPM) of Sharpe (1964) and Lintner (1965) imply that financial market participants care about risk and adjust their return expectations accordingly. Translating the latter statement into a standard dynamic CAPM-type argument (see e.g. Glosten et al., 1993; Bollerslev et al., 2013), expected aggregate market returns, r_t, can be described as

$$\mathsf{E}_t(r_{t+1}) = \gamma\omega_t^2, \tag{1}$$

where γ can be thought of as a risk aversion parameter, which according to risk return trade-off theory is expected to be > 0, and ω_t^2 is the local variance of returns with $t = 1, 2, \ldots, T$.

Equation (1) implies that given a measure for ω_t^2, returns on the market should be predictable. To investigate the empirical validity of this implication by a statistical linear regression, the researcher needs to identify a proxy for the unobservable local return variance or market risk, ω_t^2. One approach popular in the literature is to find a set of state variables that are assumed to carry information about the unobservable risk, and hence expected returns. Typical predictor variables include the dividend to price ratio (Campbell and Shiller, 1988a; Fama and French, 1988; Cochrane, 1999), the book to market ratio (Lewellen, 1999), the price earnings ratio (Campbell and Shiller, 1988b), interest rate spreads (Fama and French, 1989), and/or the consumption level relative to income and wealth, *cay* (Lettau and Ludvigson, 2001). A second commonly relied on methodology is to model $\omega_t^2 = \mathsf{Var}_t(r_{t+1})$ explicitly, and estimate it's dynamics jointly with the predictive regression within the (G)ARCH-M framework (Engle et al., 1987; Engle and Bollerslev, 1986). The recent availability of high-frequency stock market observations has opened a third possibility to proxy for risk, by employing nonparametric techniques to construct realized variance measures (see e.g. Andersen et al., 2001).

Whichever proxy the researcher decides to chose, they all seem to share the common feature of strong time series persistence. The term spread, measured as the monthly difference between a ten-year bond yield and a short-term interest rate by Campbell and Vuolteenaho (2004) and Diebold and Li (2006), has a first-order autocorrelation well above 0.9. The same measure for the price earnings ratio is almost equal to one. Stambaugh (1999) and Lewellen (2004) discover a similarly high correlation estimate for the dividend to price ratio. The latter further reports first-order autocorrelation estimates of 0.99 for the book to market ratio and the earnings price ratio. In the second framework above, the ARCH coefficient or the sum of the ARCH and the GARCH term are typically found to be close to one (for a summary, see e.g. Bollerslev et al., 1992). Similarly, the realized variance measures exhibit strong temporal dependence (see e.g. Bollerslev et al., 2012, and references therein).

[0]An extensive list of typical predictor variables can be found in Campbell (2000).

The apparent persistence in the proxy for ω_t^2, i.e. the regressor in a predictive return regression, causes econometric problems with estimation and inference that mostly arise due to the correlation between the innovations in the predictor and returns. Firstly, ordinary least squares (OLS) estimation produces a biased and/or inconsistent slope estimate of the predictive regression. If regressors are assumed $I(0)$ with autoregressive dynamics, Stambaugh (1986, 1999) describes the small-sample bias in the OLS estimate. Successively, for instance Kothari and Shanken (1997) and Lewellen (2004) derive estimates that correct for the bias. A large stream of literature describes the regressor dynamics as local to unity (LUR) processes (see e.g. Campbell and Yogo, 2006, and Jansson and Moreira, 2006), thus violating the $I(0)$ assumption. In this setup, the OLS slope estimate has an asymptotic second order bias (Phillips and Lee, 2013). It is not obvious how to correct for the presence of this asymptotic bias since the localizing coefficient cannot be consistently estimated (Phillips, 1987). Torous and Valkanov (2000) further show that if the volatility of the regressor's innovation scaled by the prediction coefficient relative to the volatility of the return innovation decreases sufficiently fast as $T \to \infty$, i.e. at rate T^{-o} with $o > 1$, then the OLS slope estimate of the predictive regression is even inconsistent.

A related econometric problem concerns the statistical inference on the predictability of returns. Within a LUR framework the t-statistic corresponding to the null hypothesis (H_0) that the regressor contains no predictive information about returns does not converge to the usual normal asymptotic distribution. Similarly, if the regressor instead is assumed to be a fractionally integrated process, $I(d)$, Maynard and Phillips (2001) show that t-statistics have nonstandard limiting distributions. Based on the work of Campbell and Yogo (2006), Cavanagh et al. (1995), and Stock (1991), who impose the former LUR-type data generating process (DGP) on the regressor, researchers have relied on confidence intervals computed using Bonferroni bounds. Predictability tests relying on this methodology are known to be conservative. A potentially severe drawback of this approach is that the confidence intervals have zero coverage probability if the regressor is stationary, as has been recently shown by Phillips (2012). IVX filtering due to Magdalinos and Phillips (2009) (see also Phillips and Lee, 2013; Gonzalo and Pitarakis, 2012) constitutes an alternative method that resolves the econometric problems of (asymptotic) bias and nonstandard inference in predictive regressions. The underlying idea is to filter the predictor such as to remove its strong temporal dependence and use the resulting series as an instrument in an instrumental variable (IV) regression. The modified variable addition method of Breitung and Demetrescu (2013), where a redundant regressor is added to the predictive regression, is a further means to achieve standard statistical inference.

A third issue arising in predictive return regressions with persistent regressors that has received less attention is the unbalanced regression phenomenon (see e.g. Banerjee et al., 1993). The studies on predictive regressions with regressor dynamics different from $I(0)$ can be classified into two sets. The first set assumes a DGP where returns are generated as noise, that is under H_0 (see e.g. Maynard and Phillips, 2001). In this setup returns are $I(0)$, whereas regressors are not, making a predictive regression unbalanced *in theory*. The second set of studies (see e.g. Torous and Valkanov (2000)) imposes a return DGP under the alternative hypothesis of predictability (H_1). In this case returns inherit the persistence of the regressor, and hence are not $I(0)$. The predictive regression is balanced *in theory*. Yet, this implications stands in stark contrast to both, economic and financial models of expected returns as well as ample empirical evidence that returns are $I(0)$ processes. It follows that predictive regressions in these frameworks are likely to be unbalanced *in practice*. The alternative DGP of Phillips and Lee (2013) that they present in the appendix is one notable exception. Small (or local) deviations from the null hypothesis are explicitly allowed for while preserving regression balancedness. Another exception is given in Maynard et al. (2013), who assume a DGP where returns are linearly related to the fractional difference of the regressor rendering returns $I(0)$.

Our work addresses all three econometric issues, that is bias/consistency, statistical inference, and regression balancedness. We cast our approach in the fractionally integrated modeling framework. There is substantial evidence that observed proxies for risk can be described as $I(d)$ processes, thus possessing long memory. For daily and weekly NASDAQ data on the log price dividend ratio, Cuñado et al. (2005) find an estimate of $d \approx 0.5$. Instead of relying on (G)ARCH models to describe $\omega_t^2 = \mathsf{Var}_t(r_{t+1})$, Baillie et al. (1996) suggest using a fractionally integrated GARCH (FIGARCH) model and find that d is larger than zero but smaller than one for the conditional exchange rate volatility. Similarly, it is well documented that realized variance measures can be modeled as fractionally integrated processes (see, among others, Ding et al. (1993), Baillie et al. (1996), Andersen and Bollerslev (1997), Comte and Renault (1998), Bollerslev et al. (2013)). Motivated by these empirical regularities, we suggest a DGP that linearly relates returns to a latent $I(0)$ predictor, ω_t^2. However, the observed regressor is corrupted by an additive long memory component. Such a DGP can be justified by the aggregation idea of Granger (1980) or the presence of structural breaks. Our approach archives balancedness under both hypothesis, the presence as well as the absence of predictability, yet a linear regression of returns on the observed regressor remains unbalanced. We show that in this case the OLS estimate in inconsistent, but standard statistical inference based on the t-statistic can be conducted. To cope with the inconsistency we successively suggest to rely on IV estimation, where the instruments are $I(0)$ and related to the unobservable risk ω_t^2. The IV estimate is consistent and the corresponding t-statistic is normally distributed. Finally, we discuss methods to establish the validity and the relevance of the instruments.

In our empirical application we demonstrate that our methodology can be used to evaluate intraday return predictability using realized and options-implied variances. We identify two instruments that are closely related to the variance risk premium and the jump component of the stock price process. We find empirical evidence that the latter two are valid and relevant instruments for the options-implied variance of the S&P 500. The IV regression of returns on this proxy for risk results in a positive and significant predictability, providing evidence for a positive risk return trade-off.

2 DGP and the Unbalanced Predictive Regression

We propose a simple framework that allows for a balanced DGP of the prediction target under the null and the alternative hypothesis, while retaining the problem of regression unbalancedness of the type $I(0)/I(d)$ in the empirical prediction model. We assume that the DGP of the true predictor variable, x_t^*, is $I(0)$. Throughout the remainder of this work, we assume that x_t^* is unobserved or latent. Further we assume that there is a function of the true predictor, $x_t = f(x_t^*)$, that is observable. Yet, this variable is corrupted by a fractionally integrated noise, which implies that the observed x_t is $I(d)$. The target, y_t, typically thought of being returns of a risky financial asset, is generated as an $I(0)$ predictive function of x_t^* with prediction coefficient β and level α, such that $\mathsf{E}_t(y_{t+1}) = \alpha + \beta x_t^*$. Equations (2)-(5) detail the assumed DGP.

$$x_t^* = \varepsilon_t \tag{2}$$

$$x_t = x_t^* + z_t \tag{3}$$

$$y_t = \alpha + \beta x_{t-1}^* + \xi_t \tag{4}$$

$$z_t = (1-L)^{-d} \eta_t, \tag{5}$$

where ε_t is independently and identically distributed (i.i.d.) with mean zero and variance σ_ε^2, and z_t is stationary fractionally integrated process with $0 < d < 1/2$, such that $(1-L)^d z_t = \eta_t$. L is the usual lag operator and $\eta_t \sim$ i.i.d. $(0, \sigma_\eta^2)$. The variance of z_t is $\sigma_z^2 = \sigma_\eta^2 \frac{\Gamma(1-2d)}{(\Gamma(1-d))^2}$. Finally, $\xi_t \sim$ i.i.d. $(0, \sigma_\xi^2)$.

Much of the existing work in the field of predictive regressions (see references in Section 1) imposes the assumption that the true predictor, x_t^*, and the observable predictor, x_t, are the same or perfectly correlated. In view of Equation (1) this would imply that market risk, ω_t^2, were observable. A very different model is considered by Ferson et al. (2003) and Deng (2014). They demonstrate the risk of spurious inference in predictive regressions, where the expected (demeaned) return βx_t^* is assumed to be independent of x_t. Note that both setups can be viewed as extremes of our DGP, where the first scenario arises if $\sigma_\eta^2 = 0$, and the second scenario occurs if $\beta = 0$ and/or $\sigma_\varepsilon^2 = 0$. Instead of imposing these extreme setups, we consider the predictor in our model to be *imperfect*. Similarly to Pastor and Stambaugh (2009) and Binsbergen and Koijen (2010) we assume that the observed variable x_t contains relevant information about the expected return, but it is imperfectly correlated with the latter.

How can we motivate the assumption that observed regressors are corrupted measures of expected returns, or more precisely that the long-memory component is viewed as noise and enters additively to the true signal? Such a DGP can be justified by the aggregation idea of Granger (1980). Assume that the observed variable x_t in (3) is composed of an aggregation of micro units $x_{i,t}$. The predictive regression for returns is typically evaluated for indices, that is an aggregation of several assets, where the predictor variable would for instance be the dividend to price ratio of an index, the conditional volatility of an index, etc. All of these processes can be viewed as examples of aggregation. Assume that $x_{i,t}$ follows a

DGP given by

$$x_{i,t} = \phi_i x_{i,t-1} + \vartheta_i w_t + \zeta_{i,t} \tag{6}$$

where w_t and $\zeta_{i,t}$ are independent $\forall\ i$. $\zeta_{i,t}$ are white noise with variance ς_i^2. In addition, there is no feedback in the system, i.e. $x_{i,t}$ does not cause w_t. Thus, $x_{i,t}$ can be viewed as $i = 1, 2, \ldots, N$ micro units of a process that are driven by their own past realizations, a common component, w_t, and an idiosyncratic shock, $\zeta_{i,t}$.

Further assume that the parameters ϕ, ϑ, and ς^2, are drawn from independent populations, and that $\phi \in (0, 1)$ is distributed as[1]

$$\mathrm{d}F(\phi^2) = \frac{2}{B(p,l)} \phi^{2p-1} \left(1 - \phi^2\right)^{l-1} \mathrm{d}\phi^2 \qquad p, l > 1, \tag{7}$$

where $B(\cdot,\cdot)$ denotes the beta function. If we sum the micro units, $x_{i,t}$, we obtain

$$x_t = \sum_{i=1}^{N} \vartheta_i \sum_{j=0}^{\infty} \phi_i^j w_{t-j} + \sum_{i=1}^{N} \sum_{j=0}^{\infty} \phi_i^j \zeta_{i,t-j}, \tag{8}$$

where $x_t = \sum_{i=1}^{N} x_{i,t}$. Granger (1980) shows that $x_t \sim I(\delta_x)$, with $\delta_x = \max(1 - l + \delta_w, 1 - l/2)$, where $w_t \sim I(\delta_w)$. Hence, if we assume that $l = 2(1-d)$ and $\delta_w = 1 - 2d$, then x_t will be integrated of the order d, i.e. $\delta_x = d \in (0, \frac{1}{2})$. Furthermore, x_t is generated by two components; the first element is a function of the common component w_t, which will be integrated of the order zero. This can be compared to the variable x_t^* in (3). The second component is a function of the idiosyncratic error terms $\zeta_{i,t}$, which will be integrated of the order d. This second component can be compared to our variable z_t in the DGP of x_t in (3). Obviously, in comparison our framework (2)-(5) is slightly less general, as we make the additional assumption that the innovations of x_t^* and z_t are i.i.d.

A different way to motivate our DGP for the observable x_t is to think of it as the sum of an expected and an unexpected component. The expected component is correctly centered at the true signal x_t^*. The unexpected component is driven by a process that has (unpredictable) breaks in the level, z_t. The argument that the persistence in observed risk measures may be due to changes in the mean is not new in the literature. For instance, Lettau and van Nieuwerburgh (2008) provide evidence for such structural level changes in the dividend to price ratio, the earning to price ratio, and the book to market ratio. They argue that these patterns could arise as a result of permanent technological innovations that affect the steady-state growth rate of economic fundamentals.

To demonstrate how unexpected structural level breaks can generate $I(d)$ dynamics in z_t, we adopt the framework of Diebold and Inoue (2001). Let s_t be a two-state Markov chain, i.e. a random variable

[1] See Beran et al. (2013), pp. 85-86.

that can assume values 1 or 2. s_t is independent of x_t^*. Define

$$\mathcal{P} = \begin{pmatrix} P\{s_t = 1|s_{t-1} = 1\} & P\{s_t = 1|s_{t-1} = 2\} \\ P\{s_t = 2|s_{t-1} = 1\} & P\{s_t = 2|s_{t-1} = 2\} \end{pmatrix} = \begin{pmatrix} \mathcal{P}_{1,1} & 1 - \mathcal{P}_{2,2} \\ 1 - \mathcal{P}_{1,1} & \mathcal{P}_{2,2} \end{pmatrix}. \tag{9}$$

Further assume that ϵ_t is a vector of size (2×1), given by

$$\epsilon_t = \begin{cases} (1 \quad 0)' & \text{if } s_t = 1 \\ (0 \quad 1)' & \text{if } s_t = 2 \end{cases}. \tag{10}$$

Now let $z_t = (\varrho_1, \ \varrho_2)' \epsilon_t$, $\varrho_1 \neq \varrho_2$. That is z_t is a variable that either has level ϱ_1 or ϱ_2, depending on the realization of the Markov chain. We assume that $\mathcal{P}_{1,1} = 1 - c_1 T^{-\delta_1}$, $\mathcal{P}_{2,2} = 1 - c_2 T^{-\delta_2}$, $\delta_1, \delta_2 > 0$, and $c_1, c_2 \in (0, 1)$, and w.l.o.g. that $\delta_1 \geq \delta_2$. If it holds that $\delta_1 < 2\delta_2 < 2 + \delta_1$, then it follows by Diebold and Inoue (2001) that $z_t \sim I(d)$, where $d = \delta_2 - \frac{1}{2}\delta_1$ and $d \in (0, \frac{1}{2})$. In addition, if the parameters satisfy the restriction that $\varrho_1 = -\varrho_2 \frac{c_1}{c_2} T^{\delta_2 - \delta_1}$ then the unconditional mean of z_t, given by[2]

$$\mathsf{E}(z_t) = \frac{\varrho_1 (1 - \mathcal{P}_{2,2}) + \varrho_2 (1 - \mathcal{P}_{1,1})}{2 - \mathcal{P}_{1,1} - \mathcal{P}_{2,2}}, \tag{11}$$

is equal to zero. This is in line with our proposed DGP of z_t in (5). As before, our DGP (2)-(5) is marginally less general. We impose that z_t is a fractional noise, whereas the resulting z_t from the regime switching framework above could have more general $I(d)$ dynamics.

To summarize, our proposed DGP (2)-(5) is consistent with the assumption of *imperfect* predictors. The imperfection is due to an $I(d)$ noise term that corrupts the true signal. This is in line with either viewing the observed predictor as a aggregation of micro units, or assuming that there are unexpected breaks in its level. Our framework further is consistent with the implication of economic/financial models and the empirical evidence that returns are $I(0)$. The DGP also incorporates the possibility of return predictability, which is justified by financial models such as (1). Finally, our setup allows for strongly persistent observed financial risk factors, which is in line with much of the empirical evidence.

Evaluating the predictability of y_t, the correct regression to estimate would be to regress y_t on x_{t-1}^*. Yet, x_{t-1}^* is not observed by the researcher, but x_t is not latent. We assume that the researcher runs the following regression misspecified and unbalanced regression

$$y_t = a + bx_{t-1} + e_t. \tag{12}$$

This motivates a further feature of our model (2)-(5). It is a stylized empirical fact that the residuals of (12) and the residuals of a time-series model for the predictor are correlated. Consider for instance the regression of stock returns on the dividend to price ratio and an autoregressive model of order one,

[2] See e.g. Hamilton, 1994, p. 684.

AR(1). The residuals of the former and the latter typically exhibit a strong negative correlation. Our DGP naturally incorporates this property. To see this, we re-write the DGP of y_t in (4) as

$$y_t = \alpha + \beta x_{t-1} + (-\beta z_{t-1} + \xi_t) . \tag{13}$$

Given our DGP, it follows that the regression residuals of (12) are composed of two elements, that is $e_t = -\beta z_{t-1} + \xi_t$. Thus, e_t will be naturally correlated with the innovation in x_t. More precisely, the covariance between the two error terms is given by

$$\mathsf{Cov}(e_t, z_t) = -\beta \sigma_z^2 \frac{d}{1-d}. \tag{14}$$

The covariance (14) is different from zero, as long as the alternative hypothesis holds, i.e. x^*_{t-1} predicts y_t with $\beta \neq 0$, the long-memory noise term is not constant, i.e. $\sigma_\eta^2 \neq 0$, and $d \in (0, \frac{1}{2})$.

3 Ordinary Least Squares Estimation

We describe the implications of regression unbalancedness and endogeneity, where the latter is caused by the correlation between the innovations in the observed noisy regressor and the target, on the OLS estimation and inference. Define two matrices $\mathbf{X}_{-1}$ and $\mathbf{y}$ of size $(T-1) \times 2$ and $(T-1) \times 1$, respectively by

$$\mathbf{X}_{-1} \equiv \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_{T-1} \end{pmatrix}' \tag{15}$$

$$\mathbf{y} \equiv \begin{pmatrix} y_2 & y_3 & \dots & y_T \end{pmatrix}'. \tag{16}$$

Theorem 1 summarizes our results for both hypotheses, the presence and absence of return predictability from x^*_{t-1}.

Theorem 1. *Let x^*_t, x_t, and y_t be generated by* (2), (3), *and* (4), *respectively. Estimate regression* (12) *by OLS, resulting in*

$$\hat{\mathfrak{b}}_{OLS} \equiv \left(\hat{a},\, \hat{b}\right)' = \left(\mathbf{X}'_{-1}\mathbf{X}_{-1}\right)^{-1} \left(\mathbf{X}'_{-1}\mathbf{y}\right). \tag{17}$$

Let $\xrightarrow{P}$ denote convergence in probability, and $\xrightarrow{D}$ convergence in distribution. Then, as $T \to \infty$:

1. *If $\beta = 0$*

$$\hat{a} \xrightarrow{P} \alpha \qquad\qquad T^{1/2}\hat{b} \xrightarrow{D} \mathcal{N}\left(0, \frac{\sigma_\xi^2}{\sigma_\varepsilon^2 + \sigma_z^2}\right)$$

$$T^{-1/2} t_a \xrightarrow{P} \frac{\alpha}{\sigma_\xi} \qquad\qquad t_b \xrightarrow{D} \mathcal{N}(0,1) .$$

$t_a = \hat{a}/\sqrt{\mathsf{Var}(\hat{a})}$ *and* $t_b = \hat{b}/\sqrt{\mathsf{Var}(\hat{b})}$ *denote the t-statistics associated with $\hat{a}$ and $\hat{b}$, respectively, and*

$\mathcal{N}(\cdot,\cdot)$ is the normal distribution. In addition, it holds that $s^2 \xrightarrow{P} \sigma_\xi^2$, where $s^2 = (T-3)^{-1}\sum_{t=2}^{T}\hat{e}_t^2$ is the variance of the OLS residuals.

2. *If $\beta \neq 0$*

$$\hat{a} \xrightarrow{P} \alpha \qquad \hat{b} \xrightarrow{P} \beta\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+\sigma_z^2}$$

$$T^{-1/2}t_a \xrightarrow{P} \frac{\alpha}{\left(\sigma_\xi^2+\beta^2\frac{\sigma_\varepsilon^2\sigma_z^2}{\sigma_\varepsilon^2+\sigma_z^2}\right)^{1/2}} \qquad T^{-1/2}t_b \xrightarrow{P} \frac{\beta\sigma_\varepsilon^2}{\left(\beta^2\sigma_\varepsilon^2\sigma_z^2+\sigma_\xi^2(\sigma_\varepsilon^2+\sigma_z^2)\right)^{1/2}},$$

where $s^2 \xrightarrow{P} \sigma_\xi^2+\beta^2\frac{\sigma_\varepsilon^2\sigma_z^2}{\sigma_\varepsilon^2+\sigma_z^2}$.

A proof of Theorem 1 can be found in Appendix B. The first part of the theorem summarizes the case in which the researcher estimates a predictive regression for unrelated variables in an unbalanced regression framework. In this situation, the OLS slope estimate $\hat{b}$ correctly converges to zero and to a normal distribution at the usual rate $T^{-1/2}$. Figure 1(i) compares the empirical distribution of $\hat{b}$ from 200,000 simulations with continuous uniformly distributed errors to the theoretical asymptotic distribution from Theorem 1. Even for small samples of size $T = 250$, the former closely approximates the latter.

In the second part of Theorem 1 we derive the asymptotic inference for the unbalanced regression framework under the alternative hypothesis that there is predictability from x_{t-1}^* on y_t. In this case, OLS produces an inconsistent estimate for β. Table 1 summarizes the simulated small sample behavior of the relative bias $\hat{b}/\beta$, with errors drawn from t-distributions. These values range from 0.17 to 0.69, which implies a substantial bias towards zero of the OLS slope estimate. The table also demonstrates that the bias is not merely present in small samples, as often the relative bias with $T = 1,000$ is larger than or equal to the corresponding value with $T = 250$, all else equal. Finally, Table 1 shows that $\hat{b}/\beta$ is independent of σ_ξ and β, but it decreases with increasing d and σ_η, and increases with increasing σ_ε. This is fully in line with the theoretical results in Theorem 1. Figure 2(i) plots the empirical average value of $\hat{b}$ for different sample sizes, T, from 200,000 simulations of the DGP (2)-(5) with t-distributed errors, proving graphical support for the analytical results in the theorem. Taken together, this implies that a non-zero linear relation between the dependent and the independent variable cannot be consistently estimated by OLS.

The results reveal that the OLS estimate has an asymptotic bias towards zero, which implies that the researcher would underestimate the implied predictive power from x_{t-1}^* on y_t. This finding stands in contrast to the conclusions in Stambaugh (1986, 1999) and Lewellen (2004). Assuming that the covariance between the prediction-regression residuals and the innovations in the predictor is negative, the latter conclude that there is a positive finite-sample bias in the OLS prediction estimate stemming from the endogeneity. Hence, if there is positive predictability the researcher will overestimate its magnitude. The problem is somewhat more severe in our setup, as $\hat{b}$ may not merely suffer from a bias, but rather is an inconsistent estimate. Given our assumptions, Regression (12) is unbalanced in addition to being endogenous. The dependent variable is $I(0)$, whereas the independent variable exhibits long memory, $I(d)$. The OLS approach attempts to minimize the sum of squared residuals in the misspecified Regression (12). This can be achieved by eliminating the memory in e_t, i.e. by letting $\hat{b} \to 0$. This finding is consistent with Maynard and Phillips (2001).

The t-statistic associated with $\hat{b}$ converges asymptotically to a standard normal limiting distribution that is free of nuisance parameters under the null hypothesis that $\beta = 0$. Small sample simulations with t-distributed errors in Table 1 support this conclusion. The size of a simple t-test on the parameter is always very close to the nominal size of 5%. Figure 1(ii) shows that even for small sample sizes the t-statistic approximates the asymptotic distribution closely. Under the alternative hypothesis, the t-statistic t_b diverges asymptotically at rate $T^{1/2}$. Figure 2(ii) supports this conclusion from Theorem 1, plotting the empirical average value of $T^{-1/2}t_b$ for different sample sizes, T, from 200,000 simulations of the DGP (2)-(5) with t-distributed errors. The implication of these results is that one can draw valid statistical inference on the significance of β. A t-test has sufficient asymptotic power to reject the null hypothesis. In other words, with T sufficiently large the researcher would eventually reject the hypothesis that the parameter is not equal to zero. The latter result makes clear that even in the unbalanced and misspecified regression framework considered here the t-statistic can be considered a useful tool to draw inference on the significance of the predictability of y_t from a latent x^*_{t-1}. Table 1 provides small sample simulation evidence for this conclusion. Drawing DGP errors from a t-distribution, we find that a t-test generally has good power. Exception from this happen mostly for small sample sizes, $T = 250$, a small absolute value of β, and large d. The worst case scenario occurs when $\sigma_\xi = \sigma_\eta = 1.73$, $\sigma_\varepsilon = 1.13$, $d = 0.49$, and $T = 250$. This is not surprising, as in this case the signal-to-noise ratio of the predictor, $\mathcal{S} \equiv \sigma_\varepsilon/\sigma_z$, is equal to 0.1615, and hence rather small. In addition, the relation between y_t and x^*_{t-1} is blurred by a noise term, ξ_t, that is more volatile than the predictor itself. All else equal, the power increases in $|\beta|$, in σ_ε, and in T; it decreases in d, σ_ξ, and σ_η.

The finding that statistical inference in our unbalanced and endogenous regression framework is not spurious may be somewhat surprising. Generally, these two phenomena when occurring jointly imply a nonstandard limiting distribution of the t-statistic under the null hypothesis. For fractionally integrated regressors, this result can be found in Maynard and Phillips (2001); the case of LUR regressors is derived in Cavanagh et al. (1995). Note, however, that given our DGP (2)-(5) the regressor is no longer endogenous under the assumption that $\beta = 0$, that is $\mathsf{Cov}(e_t, z_t) = 0$. From the literature focusing on the traditional $I(0)/I(1)$ unbalanced regression setup with exogenous regressors and i.i.d. innovations we know that the t-statistic is well behaved and converges to a standard normal random variable, as shown in Noriega and Ventosa-Santaulària (2007) and successively in Stewart (2011). Theorem 1 proves that the same result holds true in our $I(0)/I(d)$ specification.

A further implication of Theorem 1 is that the level of the conditional mean of y_t, α, can be consistently estimated by the OLS estimate $\hat{a}$, independently of the true value of β. It's associated t-statistic t_a diverges at rate $T^{1/2}$. Thus, asymptotically the researcher would correctly reject the null hypothesis that $\alpha = 0$ when the null hypothesis is false, based on a simple t-test.

To summarize, the t-statistic corresponding to an OLS estimate represents a means to identify the non-existence of a linear relationship between a random variable and its lagged latent predictor. Yet, in the present $I(0)/I(d)$ setup with unobserved regressors OLS yields an inconsistent estimate of such a linear relationship. To cope with the problem of unbalanced regressions, Maynard et al. (2013) suggest to fractionally filter the regressor; fractional differencing has also been applied by Christensen and Nielsen (2007). In this paper, we opt for a different solution to cope with the problem for several reasons. Firstly, the application of the fractional filter to the predictor requires the knowledge of d. As d in not know a priori, the researcher has to estimate it, which introduces an additional degree of uncertainty. Secondly, fractionally differencing the regressor is only a useful approach if the assumed DGP for y_t follows:

$$y_t = \alpha + \beta \left(1 - L\right)^d \tilde{x}_{t-1} + \xi_t, \tag{18}$$

with $\tilde{x}_t$ being a pure fractionally integrated process. We argue that it is difficult to justify a DGP as (18) from an economic and financial viewpoint. In a traditional $I(0)/I(1)$ framework, i.e. $d = 1$ in (18), the filter $(1 - L)^d$ applied to $\tilde{x}_{t-1}$ would imply that y_t is driven by the short-run changes of lagged $\tilde{x}_t$, instead of by its level. In the fractionally integrated setup with $d \in (0, \frac{1}{2})$, y_t in (18) would be determined by a "hybrid" of levels and changes in the predictor. This is not in line with many economic-financial models. Let y_t be the continuously compounded return on a risky financial asset, or the logarithmic dividend growth. For instance, the Dynamic Gordon Growth Model states that under rational expectations the logarithmic dividend to price ratio in *levels* should have predictive ability for future returns and/or dividend growth. It follows, that the true predictor, $\tilde{x}_{t-1}$, cannot be a fractional difference. Similarly, assume y_t is the change in the foreign exchange spot rate and let the predictor be the forward premium. The Forward Rate Unbiasedness theory implies that the expected changed in spot rates is linearly related to the *level* of the forward premium.

Finally, in our assumed DGP y_t is related to the level of a lagged latent x_t^*, which is corrupted by a persistent error. Fractional differencing in this setup cannot help solving the unbalanced regression problem. Even if d were known, filtering the observed x_{t-1} by $(1 - L)^d$ would imply an over-differencing of the true signal x_{t-1}^*. This would suggest that y_t were driven by an anti-persistent predictor.

4 Instrumental Variable Estimation

To alleviate all of the above concerns, we instead propose to estimate the linear relationship by an instrumental variable (IV) approach. Assume that the researcher has access to a valid and relevant $I(0)$

instrument, i.e. a variable that is strongly correlated with x^*_{t-1} but not with the fractional noise, z_{t-1}, and the innovation ξ_t[3]. Theorem 2 summarizes the asymptotic properties of an IV estimation of Equation 12.

Theorem 2. *Let x^*_t, x_t, and y_t be generated by* (2), (3), *and* (4), *respectively. Assume there exist K instruments*

$$q_{k,t} = \rho_k x^*_t + \upsilon_{k,t}, \quad k = 1, 2, \ldots, K, \tag{19}$$

where $\upsilon_{k,t} \sim i.i.d.\ (0, \sigma^2_{\upsilon_k})$ and $\rho_k \neq 0\ \forall k$. Further define

$$\mathbf{Q}_{-1} \equiv \begin{pmatrix} 1 & 1 & \ldots & 1 \\ q_{1,1} & q_{1,2} & \cdots & q_{1,T-1} \\ q_{2,1} & q_{2,2} & \cdots & q_{2,T-1} \\ \vdots & \ddots & \ddots & \vdots \\ q_{K,1} & q_{K,2} & \cdots & q_{K,T-1} \end{pmatrix}'. \tag{20}$$

Estimate regression (12) *by IV using $q_{k,t}$ as instruments for x_t. The IV estimate is given by*

$$\hat{\mathfrak{b}}_{IV} \equiv \left(\hat{a},\ \hat{b}\right)' = \left(\mathbf{X}'_{-1}\mathbf{Q}_{-1}\left[\mathbf{Q}'_{-1}\mathbf{Q}_{-1}\right]^{-1}\mathbf{Q}'_{-1}\mathbf{X}_{-1}\right)^{-1}\left(\mathbf{X}'_{-1}\mathbf{Q}_{-1}\left[\mathbf{Q}'_{-1}\mathbf{Q}_{-1}\right]^{-1}\mathbf{Q}'_{-1}\mathbf{y}\right). \tag{21}$$

Then, as $T \to \infty$:

1. *If $\beta = 0$*

$$\hat{a} \xrightarrow{P} \alpha \qquad T^{1/2}\hat{b} \xrightarrow{D} \mathcal{N}\left(0, \frac{\sigma^2_\xi\left(\sigma^2_\varepsilon \sum_{k=1}^K \frac{\rho_k^2}{\sigma^2_{\upsilon_k}} + 1\right)}{\sigma^4_\varepsilon \sum_{k=1}^K \frac{\rho_k^2}{\sigma^2_{\upsilon_k}}}\right)$$

$$T^{-1/2}t_a \xrightarrow{P} \frac{\alpha}{\sigma_\xi} \qquad t_b \xrightarrow{D} \mathcal{N}(0,1),$$

where $s^2 \xrightarrow{P} \sigma^2_\xi$.

2. *If $\beta \neq 0$*

$$\hat{a} \xrightarrow{P} \alpha \qquad \hat{b} \xrightarrow{P} \beta$$

$$T^{-1/2}t_a \xrightarrow{P} \frac{\alpha}{\left(\sigma^2_\xi + \beta^2\sigma^2_z\right)^{1/2}} \qquad T^{-1/2}t_b \xrightarrow{P} \beta\left(\frac{\sigma^4_\varepsilon \sum_{k=1}^K \frac{\rho_k^2}{\sigma^2_{\upsilon_k}}}{\left(\sigma^2_\xi + \beta^2\sigma^2_z\right)\left(\sigma^2_\varepsilon \sum_{k=1}^K \frac{\rho_k^2}{\sigma^2_{\upsilon_k}} + 1\right)}\right)^{1/2},$$

where $s^2 \xrightarrow{P} \sigma^2_\xi + \beta^2\sigma^2_z$.

[3]Notice that by Equation (13) it must hold that an instrument this is neither correlated with z_{t-1} nor with ξ_t will by definition also be unrelated to the error term of the unbalanced regression (12), e_t.

Theorem 2 shows that in the absence of predictability, the IV estimate $\hat{b}$ converges to a normal distribution with zero mean at the standard rate $T^{-1/2}$. Figure 3(i) shows that even if $T = 250$, $\hat{b}$ approaches the theoretical asymptotic distribution. More importantly, Theorem 2 demonstrates that IV estimation results in a consistent estimate for β. Hence, under the maintained assumption that the DGP follows (2)-(5), the predictive power of a latent variable x^*_{t-1} on y_t can be correctly inferred if the researcher finds a relevant and valid instrument for the former. Figure 4(i) supports this conclusion, plotting the average IV estimate $\hat{b}$ over 200,000 simulations for increasing T. The simulation results in Table 2 further show that the relative bias, $\hat{b}/\beta$, is very close to one even for small to moderate sample sizes. Across the set of chosen parameter values, the relative bias is bound between 1 and 1.05, when simulated errors are drawn from standard normal distributions. Finally, α can be consistently estimated by IV, as Theorem 2 proves. The proof of Theorem 2 can be found in Appendix C.

Theorem 2 further implies that the statistical significance of β can be correctly inferred from a simple t-test. Under the null hypothesis that $H_0 : \beta = 0$ the t-statistic of the IV estimate $\hat{b}$, t_b, converges to a standard normal distribution, as the simulations in Figure 3(ii) confirm. Table 2 summarizes the small sample behavior of a t-test. Overall, the size of the test is close to the nominal level of 5%, but on average the test seems somewhat undersized. Nevertheless, the size approaches 5% as T increases, as $|\rho|$ and hence $|\mathsf{Corr}(q_{k,t}, x^*_t)|$ increases, and as σ_ε increases, all else equal. We find the lowest nominal size of approximately 3% for the scenario where $\sigma_\xi = \sigma_\eta = \sigma_\upsilon = 1.73$, $\sigma_\varepsilon = 1.13$, $d = 0.49$, and $T = 250$. As mentioned in Section 3, this is to be expected as in this case $\mathcal{S}$ is small.

Contrasting the size of the t-test on the significance of β of the IV estimator in Table 2 with the corresponding size for the OLS estimator in Table 1, we can thus observe that the overall size of the former is smaller than the latter. This does not come as a surprise, as the IV estimator is generally less efficient than the OLS. Standard errors of the former are comparably slightly larger, leading to a small underrejection of the null hypothesis. It should be noted that, in our setup, there seems to be no risk of detecting predictability too often. This stands in contrast to the usual worry in the literature that predictability tests with persistent regressors may be (heavily) oversized, as pointed out by Elliott and Stock (1994) and Campbell and Yogo (2006) among others, or may even lead to spurious conclusions (Ferson et al., 2003).

Assuming that the variables follow our proposed DGP (2)-(5), we conclude that size is not an issue in our setup. Yet, the power of the OLS t-test on the significance of β in the previous section may be insufficient, especially when T is small and d is large. OLS hence implies some risk of the researcher not detecting predictability when it is present. Estimation by IV also alleviates this concern. The power of the t-test is very close to 100% across the scenarios that we consider in the simulations in Table 2. Asymptotically, t_b diverges at rate $T^{1/2}$ under $H_1 : \beta \neq 0$ as shown in Theorem 2; Figure 4(ii) depicts the convergence behavior of the statistic.

4.1 Instrument Relevance

As is generally the case, the instrument may not be irrelevant or too weak. To see this, let $q_{k,t} = \upsilon_{k,t}$ in Theorem 2 and estimate Regression (12) by IV using $q_{k,t}$ as instruments. Then, as $T \to \infty$, $\hat{b} = O_p(1)$.

To demonstrate that choosing an irrelevant instrument can lead to very undesirable properties of the IV estimation, we simulate an instrument as in (19) with $\rho_1 = 0$. Table 3 shows the the size, power, and relative bias of the resulting IV estimator and the corresponding significance test, when errors are drawn from continuous uniform distributions. The size of a standard t-test on the significance of the prediction coefficient is approximately zero. Similarly the power of the test is very low, ranging between 0.31% and 14.87%. The lack of power is is not a small sample problem, as our simulations show that the power uniformly decreases as T increases, suggesting that asympotitcally the probability to reject is zero. This implies that the researcher will tend to conclude that there is no predictability, independent of whether it is present or absent.

The low size and power properties signify that the t-statistic, t_b, is too small in absolute value if the instrument is irrelevant. This may be the result of a too small value of $|\hat{b}|$ and/or of a too large volatility of the estimate, $\sqrt{\mathsf{Var}(\hat{b})}$. Table 3 summarizes the relative bias, $\hat{b}/\beta$, which deviates wildly from the reference point of 1. IV estimation with an irrelavant instrument may lead to overestimation, underestimation, or even estimation with the incorrect sign. The relative bias covers a wide range, from -24.52 to 14.72, and the bias is independent of T. Hence, we cannot conclude that $|\hat{b}|$ is too small in absolute value. The low size and power of t-test is therefore mostly a result of a very high variance of the estimator.

To conclude, estimating (12) by IV with an irrelevant instrument leads to an inconsistent and inefficient estimator. To avoid such an outcome, we suggest a simple testing procedure. Assume that the researcher has identified a candidate instrument. Recall that the instrument follows the DGP given in (19), $q_{k,t} = \rho_k x_t^* + \upsilon_{k,t}$. As x_t^* is unobserved the researcher cannot simply regress the instrument on x_t^* to conduct inference on the value of ρ_k and thus on the instrument relevance. Instead, $q_{k,t}$ can be regress on the observed x_t by OLS, however. By Theorem 1 it holds that the slope coefficient of this regression is an inconsistent estimate of ρ_k, yet valid statistical inference using a t-test can be carried out. Thus, relying on a simple OLS t-test the researcher can infer whether the instrument is statistically irrelevant.

4.2 Instrument Validity

Besides being relevant, the instruments $q_{k,t-1}$ further need to be valid. For an instrument to be valid, it may not be correlated with the residuals of the IV regression of Regression (12), e_t. To summarize the consequences of IV estimation with an invalid instrument in finite samples, we simulate two types of invalid instruments. The first instrument is correlated with z_{t-1}, i.e. with the fractionally integrated noise. We draw a random series μ_{t-1} from a standard normal distribution with zero mean and variance

σ_μ^2, and construct the instrument as in (19) with

$$\upsilon_{k,t-1} = \mu_{t-1} + \kappa_k z_{t-1}. \tag{22}$$

This invalid instrument is correlated with e_t and it is integrated of the order d, $I(d)$. Henceforth, we refer to such instruments as invalid of type 1. Table 4 shows size and power properties, as well as the relative bias of an IV estimation of (12) with this instrument. The power of a t-test is close to 100% across the considered scenarios. Even when d is large, the researcher will still reject an incorrect null hypothesis in at least 98.60% of the cases. Similarly, the test is correctly sized at 5%. Thus, even when the instrument is invalid and $I(d)$, statistical inference on β using the t-test from the IV estimation can be conducted. Yet, the IV estimate $\hat{b}$ is biased towards zero. As this bias does not disapeear with increasing T, we conclude that $\hat{b}$ is inconsistent. The simluated relative bias, $\hat{b}/\beta$, ranges from 0.38 to 0.77. Thus, using such an invalid instrument with innovations given by (22) leads to the same outcome as when estimating Regression (12) by simple OLS.

In practice, it is fairly straightforward for the researcher to avoid invalid instruments of type 1, i.e. that are correlated with z_{t-1} as in (22). As they will be integrated of the order d, a simple statistical test for the presence of a fractional root can be relied on. Examples are the Lagrange-Multiplier tests of Robinson (1994) or Tanaka (1999), which test for an integration order d under the null hypothesis against the alternative of an integration order smaller (or larger) than d. The fractional Dickey-Fuller test of Dolado et al. (2002) is another possibility that is easy to implement.

We construct a second type of instrument in Table 4, which is correlated with ξ_t, and hence correlated with e_t and integrated of the order zero, $I(0)$. We call this second form of instrument invalidity type 2. The innovations of this instrument $q_{k,t-1}$ are simulated as

$$\upsilon_{k,t-1} = \mu_{t-1} + \kappa_k \xi_t. \tag{23}$$

Table 4 shows that choosing such an instrument can have very severe consequences. The size of a t-test on the significance of the prediction coefficient is approximately 100%. The power of the test is also close to 100% in most instances, yet in extreme cases it may drop down to as low as 2.12%. The researcher would therefore be tempted to reach the exact opposite conclusion than what it should be. If there is no predictability, one will always erroneously conclude that there is. Conversely, if there is very strong predictability, i.e. β is bounded far away from zero, we may fail to reject $\beta = 0$. The latter is especially true if d is large, T is small, σ_ξ is small, σ_η is big.

If the invalid instrument has innovations given by (23), the estimation of (12) by IV further is strongly inconsistent. The relative bias in Table 4, $\hat{b}/\beta$, shows that when $\beta = -2$, $\hat{b}$ is negative but it strongly underestimates the magnitude. When $\beta = 3$, $\hat{b}$ is positive yet it overestimates the magnitude. Finally, when $\beta = 0$[4], the estimate $\hat{b}$ is positive. We conclude that $\hat{b}$ in this case has a significant positive bias; as it does not decrease as T increases, the estimate in inconsistent.

In practice, using an invalid instrument of type 2 should be avoided at all costs. A common approach to test for the validity of an instrument is to rely on Sargan's $\mathcal{J}$ test (Sargan, 1958). Corollary 1 summarizes the asymptotic behavior of the $\mathcal{J}$ test for our DGP.

Corollary 1. *Let x_t^*, x_t, and y_t be generated by* (2), (3), *and* (4), *respectively. Assume there exist K instruments, generated by* (19)*. Estimate the following second-stage regression by OLS*

$$\hat{\mathbf{e}} = \mathbf{Q}_{-1}\varpi + \mathbf{v}, \tag{24}$$

where $\hat{\mathbf{e}}$ are the regression residuals from regression (12) *by IV. ϖ is a $(K+1)$ OLS coefficient vector and $\mathbf{v}$ is a vector of innovations. Compute the uncentered R^2 of Regression* (24) *as $R_u^2 = 1 - \frac{\hat{\mathbf{v}}'\hat{\mathbf{v}}}{\hat{\mathbf{e}}'\hat{\mathbf{e}}}$. Define a test statistic for the validity of the instruments as*

$$\mathcal{J} \equiv TR_u^2. \tag{25}$$

Then, as $T \to \infty$:

$$\mathcal{J} \xrightarrow{D} \chi^2_{(K-1)}.$$

A proof of Corollary 1 can be found in Appendix D. The corollary shows that even though the true predictor, x_{t-1}^* is not observable, we can still test whether $q_{k,t}$ is a (in)valid instrument of type 2 for the former. The statistical inference on the $\mathcal{J}$-statistic can be based on the standard χ^2 distribution. To evaluate the finite sample properties of the test, we conduct Monte Carlo experiments with 200,000 repetitions. Table 5 shows that we consider many different parameter combinations with $K = 2$, drawing innovations from t-distributions. The simulations are set in a challenging scenario, where we let $\mathsf{Corr}([q_{1,t-1},\ q_{2,t-1}]', x_{t-1}^*) = [0.85,\ 0.1]'$. Thus, there is only one strongly relevant instrument, whereas the second instrument is weakly relevant at best.

The simulation results in Table 5 suggest that the $\mathcal{J}$-test is correctly sized at a nominal level of 5%. The test is marginally oversized, with a maximal size of 5.9% across all scenarios, only when $\mathcal{S}$ is small and $\beta \neq 0$. The power of the test is fair for small values of $T = 250$, and is generally good or very good when we let T increase to 1,000. Accross all scenarios the power is substantially larger when $\beta = 0$ than when $\beta \neq 0$. This finding is not surprising, as $\hat{e}_t \to \xi_t$ when $\beta = 0$, and hence there is a very clear relation between $q_{k,t-1}$ with innovations as in (23) and $\hat{e}_t$. By the same logic, if $\beta \neq 0$ then $\hat{e}_t \to -\beta z_{t-1} + \xi_t$, and hence the signal ξ_t becomes more prominent in $\hat{e}_t$ as d decreases, σ_η decreases, and/or σ_ξ increases, thus increasing the power of the test. Finally, the power of the test decreases as we let $\mathsf{Corr}([q_{1,t-1},\ q_{2,t-1}]', \xi_t)$

[4]These estimates are not reported here to save space. The results are available from the authors upon request.

decrease from $[0.5,\ -0.6]'$ to $[-0.4,\ 0.3]'$.

We conclude that there is almost no risk of overrejecting instrument validity in finite samples when the instrument is valid of type 2. However, there is a small chance to erroneously conclude that an invalid instrument is valid, due to insufficient power in (very) small samples. To safeguard against this, we recommend that the researcher chose a conservative confidence level, for instance 10%.

5 Predicting Returns on the S&P 500

To exemplify that the suggested approach from the previous sections can help alleviate some of the concerns in empirical asset pricing, we predict daily returns, r_{t+1}, $t = 1, 2, \ldots, T$, on the S&P 500 stock market index. We consider the data period from February 2, 2000 until April 25, 2013, resulting in $T = 3325$ observations. We assume that risk or uncertainty in the financial market, i.e. ω_t^2 in (1), can be proxied by observable variance measures. Our first risk proxy is the realized return variance, $RV_{\mathrm{RL},t}$, computed on the basis of intradaily observations spaced into 5-minute intervals. Under certain regularity conditions, $RV_{\mathrm{RL},t}$ converges to the daily quadratic variation of returns, as shown by Andersen et al. (2001), Barndorff-Nielsen and Shephard (2002), and Meddahi (2002). Our second measure is the bipower variation, $BV_{\mathrm{RL},t}$, of Barndorff-Nielsen and Shephard (2004), which converges to the integrated variance of returns. The three series, r_t, $RV_{\mathrm{RL},t}$, and $BV_{\mathrm{RL},t}$, are obtained from the Oxford-Man Institute's "Realised Library"[5]. As a final proxy for ω_t^2, we consider the volatility index, $VIX_{\mathrm{CBOE},t}$. It is a measure for the risk-neutral expectation of return volatility over the next month, and as such can be viewed as the options-implied volatility. We obtain the series $VIX_{\mathrm{CBOE},t}$, which is traded on the Chicago Board of Options Exchange (CBOE), from the WRDS database. We transform the data series into monthly variance units by

$$vix_t^2 = \frac{30}{365} VIX_{\mathrm{CBOE},t}^2. \tag{26}$$

Whereas vix_t^2 is related to the return variation over a month, the raw series $RV_{\mathrm{RL},t}$ and $BV_{\mathrm{RL},t}$ measure daily variation. To align the three measures, we modify the latter two as follows.

$$rv_t = \sum_{i=1}^{22} \left(RV_{\mathrm{RL},t-i+1} \times 100^2 + \left\{ \left[\ln \frac{P_{t-i+1}^{(\mathrm{open})}}{P_{t-i}^{(\mathrm{close})}} \right] \times 100 \right\}^2 \right) \tag{27}$$

$$bv_t = \sum_{i=1}^{22} \left(BV_{\mathrm{RL},t-i+1} \times 100^2 + \left\{ \left[\ln \frac{P_{t-i+1}^{(\mathrm{open})}}{P_{t-i}^{(\mathrm{close})}} \right] \times 100 \right\}^2 \right). \tag{28}$$

It is well known that the three variance series exhibit strongly dependent dynamics that closely resemble fractionally integrated processes (see e.g. Bollerslev et al., 2013, and references therein). At the same time, asset returns, especially at the daily frequency level, are known to exhibit almost no serial correlation.

[5] Available at `http://realized.oxford-man.ox.ac.uk/`.

This renders a regression of the following type unbalanced.

$$r_{t+1} = a + bx_t + e_{t+1}, \tag{29}$$

where for the remainder of this section x_t is either rv_t, bv_t, or vix_t^2. To avoid any overlap between daily returns, r_{t+1}, and the realized variance and bipower variation measures, we define returns as intraday net returns[6]

$$r_t = \left[\frac{P_t^{(\text{close})} - P_t^{(\text{open})}}{P_t^{(\text{open})}} \right] \times 100. \tag{30}$$

To provide evidence that Regression (29) is indeed unbalanced for our data set, we estimate the respective fractional integration order, d_i, of the four series, rv_t, bv_t, vix_t^2, and r_t, jointly.

It is common to rely on semiparametric techniques for the estimation of d_i, as they permit the researcher to assess the long-memory behavior of the process close to frequency zero, while allowing for some unparameterized dynamics at intermediate or high frequencies. There are two commonly used classes of semiparametric estimators; the log-periodogram estimators introduced by Geweke and Porter-Hudak (1983) and the local Whittle estimators, originally developed by Künsch (1987). We rely on the latter class, since it is more robust and efficient, as pointed out by Henry and Zaffaroni (2002). The exact local Whittle (EW) due to Shimotsu and Phillips (2005) is particularly attractive, since it is consistent and asymptotically normally distributed for any value of d_i. Nielsen and Shimotsu (2007) derive a multivariate version of the EW, which we apply for the joint estimation of d_{rv}, d_{bv}, d_{vix^2}, and d_r[7].

Table 6 summarizes our results. The realized variance and the bipower variation are integrated of the order $I(0.35)$ and $I(0.34)$, respectively. At a 5% significance level, we reject that $d_i = 0$ and $d_i = 1$ for both series, yet we fail to reject that $d_i = 0.5$. The point estimate for the memory of the volatility index, vix_t^2, is somewhat higher, $\hat{d}_{vix^2} = 0.44$. According to the t-test of Nielsen and Shimotsu (2007) for the equality of d_i, we cannot reject that the three variance series are integrated of the same order, however. Intraday returns in turn are integrated of the approximate order zero, and we fail to reject $d_i = 0$. The t-tests for $H_0 : d_i = d_j$ indicate the we reject the hypothesis that variance series and returns are integrated of the same order, which makes Regression (29) unbalanced. For further evidence of the apparently distinct dynamics of the three variance series and stock returns, see also Figure 5, where we plot the autocorrelations of the four processes. Whereas shocks to daily returns die out immediately, shocks to rv_t, bv_t, and vix_t^2 are highly persistent. As opposed to the stationary return process, it takes many lags to revert the effect of a shock to the variance.

[6] All estimation results in this section remain virtually unchanged if we rely on daily close-to-close returns, instead. Here we only report the results for intraday returns; outcomes with close-to-close returns are available from the authors upon request.

[7] The consistency and asymptotic properties of the EW estimator rely on the knowledge of the true mean of the data generating process. As this value is not known in practical applications, we modify the EW to account for this uncertainty, relying on the two-step feasible EW estimator of Shimotsu (2010).

One shortcoming of the approach above is that the EW is not explicitly robust to the presence of additive perturbations, which are present in three variance processes, rv_t, bv_t, and vix_t^2, under the maintained assumption that x_t follows a short-memory signal plus a long-memory noise process as (3). To robustify our approach, we further rely on the trivariate version of the modified EW estimator of Sun and Phillips (2004) (TEW). Let $X_t \equiv [rv_t,\ bv_t,\ vix_t^2]'$. The underlying assumption of the TEW estimation approach in our setup is that the spectral density of X_t at frequency λ is given by

$$f_X(\lambda) \sim D\tau D' + \iota H \quad \text{as } \lambda \to 0+, \tag{31}$$

where $D = (\text{diag}[\lambda^{-d_{rv}},\ \lambda^{-d_{bv}},\ \lambda^{-d_{vix^2}}])$, and τ is a diagonal matrix with elements $f_{\eta_i}(0)$. Hence, we assume that the fractional-noise series, $z_{t,i}$, are uncorrelated across the three variance measures. H is a (3×3) matrix of ones; thus we impose that the signal $x_t^* = \omega_t^2$ is the same for rv_t, bv_t, and vix_t^2, and has variance $2\pi\iota$. We estimate the respective fractional order of integration of the three series jointly with the ratio τ/ι, by concentrated TEW-likelihood. We find that $\hat{d}_{rv}$ =0.36, $\hat{d}_{bv}$ =0.46, $\hat{d}_{vix^2}$ =0.33. The exact asymptotic properties of the TEW are unknown, yet Sun and Phillips (2004) conjecture that the distribution is normal and that standard errors are bound between [0.11,0.15]. The estimates for d_i are thus different from zero and statistically indistinguishable from the non-robust estimates in Table 6. From the point estimates for τ/ι, we can compute the implied signal-to-noise ratio; we find $\mathcal{S}_{rv}$ =50.12, $\mathcal{S}_{bv}$ =24.28, and $\mathcal{S}_{vix^2}$ =15.44. This suggests that the variation in the signal is strong relative to the volatility in the fractionally integrated noise for all three variance series[8].

Next we investigate the consequences of ignoring the regression unbalancedness and instead estimating the prediction regression by OLS. Table 7 outlines the results. If we predict daily returns on the S&P 500 by rv_t the prediction coefficient is very close to zero and it is statistically insignificant. Similarly, if we evaluate the unbalanced regression (29) with $x_t = bv_t$ we obtain a very small and insignificant slope estimate. Yet, when we use the vix_t^2 series to predict returns, we find a positive $\hat{b} = 0.15 \times 10^{-2}$ and it is significantly different from zero. The estimated coefficient is very small, however, and we know from Theorem 1 that the estimate is inconsistent.

To alleviate the problems associated with the unbalanced OLS regression, we define a set of instruments for IV estimation. To that end, note that there is substantial evidence that there is a linear long-run relation between rv_t and vix_t^2 that is $I(0)$. For instance, Bandi and Perron (2006) and Christensen and Nielsen (2006) find evidence of fractional cointegration between the two series. Furthermore, if the cointegrating vector is equal to $[-1,\ 1]'$, then the resulting cointegrating series corresponds to the monthly version of the variance risk premium, vrp_t, as defined by Bollerslev et al. (2009). The latter argue that vrp_t may be viewed as bet on pure volatility; as such it is reasonable to expect that the measure is

[8] We expect the confidence bands for the estimates for $\mathcal{S}$ to be very wide, and hence their values have to be interpreted with care and rather viewed as indicative. The reason is that the likelihood function for the TEW becomes flat in $(\tau/\iota)^{-1}$ when $T \to \infty$, as shown by Hurvich et al. (2005).

closely liked to the local variance in (1), $\omega_t^2 = x_t^*$, that we are aiming at proxying with the instrument. Bollerslev et al. (2009) and Bollerslev et al. (2013) also present evidence that vrp_t can predict aggregate market returns, which is further motivation for considering the measure to be a relevant instrument in our framework.

Besides the cointegrating relation between rv_t and vix_t^2, we expect that there is a long-run relation between rv_t and bv_t, as both series measure the monthly integrated variance of stock returns. Following the arguments in Barndorff-Nielsen and Shephard (2004), Andersen et al. (2007), and Huang and Tauchen (2005), the cointegrating relation between rv_t and bv_t represents the contribution of price jumps to the variance, if the cointgerating vector is equal to $[1,\ -1]'$. For instance, Andersen et al. (2007) find that the jump component exhibits a much lower degree of persistence than the two series rv_t and bv_t, providing evidence for a fractional cointegration relation. Jumps are closely related to stock market volatility; Corsi et al. (2010) and Andersen et al. (2007), among others, find that the former is an important predictor of the latter. Therefore we anticipate jumps to be a relevant instrument for $\omega_t^2 = x_t^*$.

We investigate the potential cointegration relation by a restricted version of the co-fractional vector autoregressive model of Johansen (2008, 2009) and Johansen and Nielsen (2012), given by

$$\Delta^d X_t = \varphi \left[\theta' \left(1 - \Delta^d \right) X_t \right] + \sum_{i=1}^{n} \Gamma_i \Delta^d \left(1 - \Delta^d \right)^i X_t + u_t. \tag{32}$$

We rely on model (32) because it allows us to identify a cointegration relation between the variables, while at the same time explicitly accounting for possible dynamics at higher frequencies, which may be present due to the overlapping nature of rv_t and bv_t[9]. Given the identification problems of the model (see, Carlini and Santucci de Magistris, 2013), we initially fix the cointegration rank $r = 2$ and estimate (32) by restricted maximum likelihood. Subsequently, we test for cointegration. For $\hat{d} = 0.38$ (SE($\hat{d}$)=0.10) and $n = 3$ we find the cointegrating matrix estimate

$$\hat{\theta}' = \begin{pmatrix} 1 & -1.1938 & 0 \\ -1.0111 & 0 & 1 \end{pmatrix}. \tag{33}$$

Johansen (2008) states that model (32) has a solution and $\theta' X_t \sim I(0)$ if the following conditions are satisfied. Firstly, r needs to be smaller than 3. The value of the likelihood-ratio (LR) statistic of Johansen and Nielsen (2012) that provides a test for $H_0 : r \leq 2$ against $r \leq 3$ is equal to 3.7709; thus we fail to reject the null hypothesis. Secondly, it must hold that $|\varphi_\perp' \left(I_{3\times3} - \sum_{i=1}^{n} \Gamma_i \right) \theta_\perp| \neq 0$. In our estimation this value is equal to -1.46, i.e. different from zero. Thirdly, the roots c of the characteristic polynomial $|(1-c)I_{3\times3} - \varphi\theta' c - (1-c)\sum_{i=1}^{n} \Gamma_i c^i| = 0$ must be either equal to one or $\notin$ a complex disk $\mathbb{C}_d$. Figure 6

[9]The Matlab code for the maximum-likelihood estimation of the parameters of model (32) has been provided by Nielsen and Morin (2012).

shows that all roots fulfill this final condition. We conclude that we have identified two instruments

$$q_t = \begin{pmatrix} q_{1,t} \\ q_{2,t} \end{pmatrix} = \hat{\theta}' X_t \tag{34}$$

that are integrated of the order zero. Hence, q_t are not invalid instruments of type 1 as described in Section 4.2, that is q_t is not correlated with the $I(d)$ noise term, z_t.

If we estimate a restricted version of our benchmark co-fractional model, where $\theta_{(2,1)} = -1$ and $\theta_{(1,2)} = -1$, we obtain a LR statistic of 6.6354. This implies that we reject the restriction. Whereas the second cointegrating relation, $q_{2,t}$, is essentially the variance risk premium of Bollerslev et al. (2009), $q_{1,t}$ differs slightly from the pure jump contribution, i.e. the squared jump sizes over one month. More precisely, $q_{1,t} \approx \sum_{i=1}^{22} \sum_{j=1}^{N_{t-i+1}} \psi_{t-i+1,j}^2 - 0.19 bv_t$, where $\psi_{t,j}$ is the size of the jth jump on day t, and N_t denotes the total number of jumps in a day. Noting that $|rv_t - bv_t| \geq |q_{1,t}|$ for more than 95% of the total observations in our sample, $q_{1,t}$ thus reduces the absolute value of the jump component by $0.19bv_t$, that is setting it closer to zero. This can be viewed as a crude approximation to the standard approach of only considering significant jumps (see, for instance, Tauchen and Zhou, 2011 and Andersen et al., 2007). Relying on the method outlined Section 4.1, we now investigate whether the two instruments are relevant. Regressing $q_{1,t}$ on rv_t, bv_t, and vix_t^2, respectively, we find the corresponding t-statistics, $t_{\hat{\rho}_1}$, to be equal to -6.54, -12.31, and -3.33. The *jump instrument* is a relevant instrument for the unobserved stationary component of all three variance series. Carrying out the same analysis for $q_{2,t}$, we find the respective values for $t_{\hat{\rho}_2}$ to be equal to -11.42, -11.98, 14.11, suggesting that also the *variance risk premium instrument* is strongly relevant.

Table 7 lists the outcomes of the IV estimations of Regression (29), using $q_{1,t}$ and $q_{2,t}$ from (34) as instruments. If we predict intraday returns with rv_t, we find a negative prediction coefficient, $\hat{b} = -0.13 \times 10^{-1}$, that is statistically significant. This finding stands in contrast to the OLS estimation result, where we discover that rv_t does not contain an $I(0)$ component that significantly predicts returns. The solution to this puzzle can be found in the the $\mathcal{J}$-test for instrument validity of type 2. The $\mathcal{J}$-statistic is equal to 13.73, which is well above the $\chi^2_{(1)}$ critical value at any commonly considered confidence level. Hence, the *jump instrument* and the *variance risk premium instrument* for the unobserved stationary component in rv_t are invalid. From the simulations in Table 4 we know that if the instrument(s) are invalid of type 2, the researcher is likely to find predictability even though there is none. This explains why we erroneously conclude that there is significant return predictability in the series rv_t from the IV estimation. Furthermore, the slope estimate $\hat{b}$ is known to have an asymptotic upward bias. To conclude, rv_t does not carry predictive information for daily returns on the S&P 500. For the bv_t series, the results in Table 7 are qualitatively the same.

Finally, we consider vix_t^2 as a predictor. The OLS estimation results imply that there is a positive predictability from vix_t^2 on r_{t+1}, but the prediction-coefficient estimate of 0.15×10^{-2} is asymptotically biased towards zero. If we instead predict r_{t+1} by vix_t^2 using the two identified instruments and IV estimation, we obtain a statistically significant slope estimate of $\hat{b} = 0.13 \times 10^{-1}$. This estimate is almost nine times larger than the corresponding inconsistent OLS estimate. The $\mathcal{J}$-statistic is equal to 1.41. As this value is smaller than the corresponding $\chi^2_{(1)}$ critical value, even if we consider a significance level of 20%, we conclude that *jump* and the *variance risk premium* are valid instruments in this case. Hence, we find strong evidence that there is an unobservable $I(0)$ component, $x_t^* = \omega_t^2$, contained in the vix_t^2 series that positively predicts future daily stock returns, but that is corrupted by a fractionally integrated noise term, z_t. The risk-return trade-off thus is positive.

6 Concluding remarks

This paper presents a novel DGP that accounts for many theoretical and empirical features of the return prediction literature, such as for instance persistence in the observed predictors and the stationary noise-type behavior of returns. Assuming that the practitioner estimates a misspecified and unbalanced predictive regression, where the regressors are imperfect measures of the true predictor variable, we show that OLS estimation of the predictive regression results in inconsistent estimates for the prediction coefficient. Nevertheless, standard statistical inference based on t-tests remains valid. To avoid the problem of obtaining an inconsistent estimate for the prediction coefficient, we propose an IV estimation method. If the practitioner has access to a valid and relevant $I(0)$ instrument, IV estimation results in a consistent estimate for the predictive coefficient and standard statistical inference on predictability can be carried out.

Our paper is closely related to the work on predictive regressions with IVX filtering of Magdalinos and Phillips (2009) and Phillips and Lee (2013), where the predictor is assumed to have LUR dynamics. Similarly to our approach, the underlying idea is to find an instrument that is less persistent that the regressor and use it in an IV regression. They show that consistency of the prediction estimate and standard statistical inference can be achieved in this framework[10], which is in line with our conclusions. From a theoretical viewpoint, Phillips and Lee (2013) also explicitely addresses the issue of an unbalanced regression and their extended framework presented in the appendix permits local deviations from H_0 while retaining balancedness. Important differences to our work are that our setup allows for unrestricted deviations for the null hypothesis of no predictability. Our theoretical predictive equation remains balanced for any value of the prediction coefficient. Secondly, whereas Phillips and Lee (2013) assume that the true predictor is observed, we view regressors as imperfect. Lastly, the instrument in Magdalinos and Phillips (2009) and Phillips and Lee (2013) is easy to find, as it is a filtered version of the predictor itself, and it is relevant and valid by definition. In our setup, the practitioner has to find an instrument and subsequently test for instrument relevance and validity. To that end, we discuss methodologies to investigate instrument relevance and validity.

[10]Note that the framework of Phillips and Lee (2013) permits multivariate regressors and discusses multi-period predictions, which we do not consider in this work.

Finally, we apply the methods outlined in this paper to the investigation of the predictabilility of daily returns on the S&P 500 stock market. Relying on an analysis of fractional cointegration, we provide one suggestion of how an $I(0)$ instrument can be identified. We find evidence of significant return predictability and a positive risk-return trade-off, using the suggested IV approach.

Appendix

A Useful Lemma

Lemma 1 will prove useful for the derivations of the results in this paper.

Lemma 1. *Let a_t and b_t be two independent processes satisfying $a_t \sim$ i.i.d. $(0, \sigma_a^2)$ and $b_t \sim I(d)$, with variance σ_b^2 and $0 < d < \frac{1}{2}$ for $t = 1, 2, \cdots, T$. Then $c_t = b_t a_t$ is a stationary martingale difference sequence (m.d.s.) with finite variance $\sigma_c^2 = \sigma_b^2 \sigma_a^2$.*

Proof: The relevant moments of c_t are given by

$$\begin{aligned}
\mathsf{E}(c_t) &= \mathsf{E}(b_t a_t) = \mathsf{E}(b_t)\mathsf{E}(a_t) = 0 && \forall t \in \{1, \ldots, T\} \\
\mathsf{E}(c_t^2) &= \mathsf{E}(b_t^2 a_t^2) = \sigma_b^2 \sigma_a^2 && \forall t \in \{1, \ldots, T\} \\
\mathsf{E}(c_t c_s) &= \mathsf{E}(b_t a_t b_s a_s) = \mathsf{E}(b_t b_s)\mathsf{E}(a_t a_s) = 0 && \forall t \neq s;\ t, s \in \{1, \ldots, T\} \\
\mathsf{E}(c_t | c_{t-1}, c_{t-2}, \ldots, c_1) &= \mathsf{E}(a_t b_t | a_{t-1} b_{t-1}, a_{t-2} b_{t-2}, \ldots, a_1 b_1) \\
&= \mathsf{E}\left(\mathsf{E}\left[a_t b_t | a_{t-1}, a_{t-2}, \ldots, a_1, b_t, b_{t-1}, b_{t-2}, \ldots, b_1\right] | a_{t-1} b_{t-1}, a_{t-2} b_{t-2}, \ldots, a_1 b_1\right) \\
&= \mathsf{E}\left(b_t \mathsf{E}\left[a_t | a_{t-1}, a_{t-2}, \ldots, a_1, b_t, b_{t-1}, b_{t-2}, \ldots, b_1\right] | a_{t-1} b_{t-1}, a_{t-2} b_{t-2}, \ldots, a_1 b_1\right) \\
&= \mathsf{E}\left(b_t \times 0 | a_{t-1} b_{t-1}, a_{t-2} b_{t-2}, \ldots, a_1 b_1\right) = 0 && \forall t \in \{2, \ldots, T\}.
\end{aligned}$$

The equalities are a natural consequence of the assumption of independence between b_t and a_t and the fact that a_t is i.i.d. It follows that the memory structure of an $I(d)$ process *disappears* when multiplied by an i.i.d. process. This is the main driver of the results in Theorem 2 and Corollary 1. As $c_t = a_t b_t$ is also ergodic, the Ergodic Stationary Martingale Difference Central Limit Theorem (ESMD-CLT) of Billingsley (1961) applies; the order of convergence of $\sum_{t=1}^{T} c_t$ is $O_p(T^{1/2})$.

B Proof of Theorem 1

B.1 If $\beta \neq 0$:

The OLS estimator of regression model (12) is given by $\hat{\mathsf{b}}_{OLS} = \left(\mathbf{X}'_{-1}\mathbf{X}_{-1}\right)^{-1}\left(\mathbf{X}'_{-1}\mathbf{y}\right)$, where

$$\begin{aligned}
\left(\mathbf{X}'_{-1}\mathbf{X}_{-1}\right)^{-1} &= \frac{1}{T\sum x_{t-1}^2 - \left(\sum x_{t-1}\right)^2} \begin{pmatrix} \sum x_{t-1}^2 & -\sum x_{t-1} \\ -\sum x_{t-1} & T \end{pmatrix}, \\
\mathbf{X}'_{-1}\mathbf{y} &= \begin{pmatrix} \sum y_t \\ \sum y_t x_{t-1} \end{pmatrix},
\end{aligned}$$

with x_{t-1} and y_t generated by Equations (3) and (4), respectively, and $\hat{\mathsf{b}}_{OLS} = \left(\hat{a},\ \hat{b}\right)'$. $\mathbf{X}_{-1}$ and $\mathbf{y}$ are defined as in Equations (15) and (16), and all sums run from $t = 1$ to T unless stated otherwise[11]. It

[11] Strictly speaking, T should be replaced by $T - 1$ in all equations, as we lose one observation by lagging x_t; similarly, all sums should run from $t = 2$ to T. Asymptotically, this will make no difference, however.

follows that

$$\hat{a} = \frac{\sum x_{t-1}^2 \sum y_t - \sum x_{t-1} \sum y_t x_{t-1}}{T \sum x_{t-1}^2 - \left(\sum x_{t-1}\right)^2}, \tag{B1}$$

$$\hat{b} = \frac{T \sum y_t x_{t-1} - \sum y_t \sum x_{t-1}}{T \sum x_{t-1}^2 - \left(\sum x_{t-1}\right)^2}. \tag{B2}$$

To derive the asymptotic behavior of the estimators (B1) and (B2), along with the associated t-statistics, it is necessary to obtain the limit expression of the sums that appear in the equations. They are summarized in Table B1, along with their respective convergence rates. All of the convergence rates (see the *underbraced* expressions) can be found in Tsay and Chung (2000) except for the normalization ratio of $\sum \varepsilon_{t-1} z_{t-1}$ and $\sum \xi_t z_{t-1}$, which follow from Lemma 1.

$\sum x_{t-1}$	$=$	$\underbrace{\sum \varepsilon_{t-1}}_{O_p(T^{1/2})} + \underbrace{\sum z_{t-1}}_{O_p(T^{d+1/2})}$
$\sum x_{t-1}^2$	$=$	$\underbrace{\sum \varepsilon_{t-1}^2}_{O_p(T)} + \underbrace{\sum z_{t-1}^2}_{O_p(T)} + 2\underbrace{\sum \varepsilon_{t-1} z_{t-1}}_{O_p(T^{1/2})}$
$\sum y_t$	$=$	$\alpha T + \beta \sum \varepsilon_{t-1} + \underbrace{\sum \xi_t}_{O_p(T^{1/2})}$
$\sum y_t^2$	$=$	$\alpha^2 T + \beta^2 \sum \varepsilon_{t-1}^2 + \underbrace{\sum \xi_t^2}_{O_p(T)} + 2\alpha\beta \sum \varepsilon_{t-1} + 2\alpha \sum \xi_t + 2\beta \underbrace{\sum \xi_t \varepsilon_{t-1}}_{O_p(T^{1/2})}$
$\sum y_t x_{t-1}$	$=$	$\alpha \sum \varepsilon_{t-1} + \alpha \sum z_{t-1} + \beta \sum \varepsilon_{t-1}^2 + \beta \sum \varepsilon_{t-1} z_{t-1} + \sum \xi_t \varepsilon_{t-1} + \underbrace{\sum \xi_t z_{t-1}}_{O_p(T^{1/2})}$

Table B1: Expressions for sums in Theorem 1.

For ease of exposition, denote $\hat{a}^{(n)}$ and $\hat{a}^{(d)}$ the numerator and denominator of $\hat{a}$, respectively, and substitute the expressions from Table B1.

$$\begin{aligned}
\hat{a}^{(n)} &= \left(\sum x_{t-1}^2\right)\left(\sum y_t\right) - \left(\sum x_{t-1}\right)\left(\sum y_t x_{t-1}\right) \\
&= \underbrace{\alpha T \sum \varepsilon_{t-1}^2 + \alpha T \sum z_{t-1}^2}_{O_p(T^2)} - \underbrace{\beta \sum \varepsilon_{t-1}^2 \sum z_{t-1}}_{O_p(T^{d+3/2})} - \underbrace{\alpha \left(\sum z_{t-1}\right)^2}_{O_p(T^{2d+1})} \\
&\quad + \underbrace{\sum \xi_t \sum \varepsilon_{t-1}^2 + \beta \sum \varepsilon_{t-1} \sum z_{t-1}^2 + \sum \xi_t \sum z_{t-1}^2 + 2\alpha T \sum \varepsilon_{t-1} z_{t-1}}_{O_p(T^{3/2})} \\
&\quad - \underbrace{2\alpha \sum \varepsilon_{t-1} \sum z_{t-1} - \beta \sum z_{t-1} \sum \varepsilon_{t-1} z_{t-1} - \sum z_{t-1} \sum \xi_t \varepsilon_{t-1} - \sum z_{t-1} \sum \xi_t z_{t-1}}_{O_p(T^{d+1})} \\
&\quad + \underbrace{\beta \sum \varepsilon_{t-1} z_{t-1} \sum \varepsilon_{t-1} + 2 \sum \varepsilon_{t-1} z_{t-1} \sum \xi_t - \sum \varepsilon_{t-1} \sum \xi_t \varepsilon_{t-1} - \sum \varepsilon_{t-1} \sum \xi_t z_{t-1} - \alpha \left(\sum \varepsilon_{t-1}\right)^2}_{O_p(T)}
\end{aligned} \tag{B3}$$

$$\begin{aligned}\hat{a}^{(d)} &= T\sum x_{t-1}^2 - \left(\sum x_{t-1}\right)^2 \\ &= \underbrace{T\sum \varepsilon_{t-1}^2 + T\sum z_{t-1}^2}_{O_p(T^2)} + \underbrace{2T\sum \varepsilon_{t-1}z_{t-1}}_{O_p(T^{3/2})} - \underbrace{\left(\sum z_{t-1}\right)^2}_{O_p(T^{2d+1})} - \underbrace{2\sum \varepsilon_{t-1}\sum z_{t-1}}_{O_p(T^{d+1})} - \underbrace{\left(\sum \varepsilon_{t-1}\right)^2}_{O_p(T)} \qquad \text{(B4)}\end{aligned}$$

It follows that the expression for $\hat{a}$ simplifies to

$$\hat{a} = \frac{\alpha T\left(\sum \varepsilon_{t-1}^2 + \sum z_{t-1}^2\right) + O_p(T^{d+3/2})}{T\left(\sum \varepsilon_{t-1}^2 + \sum z_{t-1}^2\right) + O_p(T^{3/2})}.$$

Dividing both, the numerator and the denominator, by T^2 and letting $T \to \infty$, we obtain

$$\operatorname*{plim}_{T\to\infty} \hat{a} = \alpha, \qquad \text{(B5)}$$

since the remaining terms collapse. Now let $\hat{b}^{(n)}$ and $\hat{b}^{(d)}$ be the numerator and denominator of $\hat{b}$, respectively. Then

$$\begin{aligned}\hat{b}^{(n)} &= T\sum y_t x_{t-1} - \left(\sum y_t\right)\left(\sum x_{t-1}\right) \\ &= \underbrace{\beta T\sum \varepsilon_{t-1}^2}_{O_p(T^2)} + \underbrace{\beta T\sum \varepsilon_{t-1}z_{t-1} + T\sum \xi_t\varepsilon_{t-1} + T\sum \xi_t z_{t-1}}_{O_p(T^{3/2})} \\ &\quad - \underbrace{\beta\sum \varepsilon_{t-1}\sum z_{t-1} - \sum \xi_t\sum z_{t-1}}_{O_p(T^{d+1})} - \underbrace{\beta\left(\sum \varepsilon_{t-1}\right)^2 - \sum \xi_t\sum \varepsilon_{t-1}}_{O_p(T)}. \qquad \text{(B6)}\end{aligned}$$

Noting that $\hat{b}^{(d)} = \hat{a}^{(d)}$, we obtain

$$\hat{b} = \frac{\beta T\sum \varepsilon_{t-1}^2 + O_p(T^{3/2})}{T\left(\sum \varepsilon_{t-1}^2 + \sum z_{t-1}^2\right) + O_p(T^{3/2})}.$$

Dividing $\hat{b}^{(n)}$ and $\hat{b}^{(d)}$ by T^2, in the limit we have

$$\operatorname*{plim}_{T\to\infty} \hat{b} = \beta\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + \sigma_z^2}. \qquad \text{(B7)}$$

Next, we demonstrate the derivation of the asymptotic expression for the variance of the regression residuals, s^2.

$$\begin{aligned}s^2 &= \frac{\left(\mathbf{y} - \mathbf{X}_{-1}\hat{\mathbf{b}}_{OLS}\right)'\left(\mathbf{y} - \mathbf{X}_{-1}\hat{\mathbf{b}}_{OLS}\right)}{T-2} \qquad \text{(B8)} \\ &= \frac{T\sum y_t^2\sum x_{t-1}^2 - \sum y_t^2\left(\sum x_{t-1}\right)^2 - T\left(\sum y_t x_{t-1}\right)^2 - 2\sum x_{t-1}\sum y_t x_{t-1}\sum y_t + \sum x_{t-1}^2\left(\sum y_t\right)^2}{(T-2)\left(T\sum x_{t-1}^2 - \left(\sum x_{t-1}\right)^2\right)}.\end{aligned}$$

Substituting the expressions from Table B1, we obtain the following solutions for the numerator, $s^{2(n)}$, and denominator, $s^{2(d)}$, of s^2.

$$s^{2(n)} = \underbrace{-T\sum\varepsilon_{t-1}^2\sum\xi_t^2 - T\sum z_{t-1}^2\sum\xi_t^2 - T\beta^2\sum\varepsilon_{t-1}^2\sum z_{t-1}^2}_{O_p(T^3)} + o_p(T^3) \tag{B9}$$

$$s^{2(d)} = \underbrace{-T^2\sum\varepsilon_{t-1}^2 - T^2\sum z_{t-1}^2}_{O_p(T^3)} + o_p(T^3). \tag{B10}$$

Thus, we obtain the following.

$$s^2 = \frac{\frac{1}{T^2}\left(\sum\xi_t^2\sum\varepsilon_{t-1}^2 + \sum\xi_t^2\sum z_{t-1}^2 + \beta^2\sum z_{t-1}^2\sum\varepsilon_{t-1}^2\right) + o_p(1)}{\frac{1}{T}\left(\sum\varepsilon_{t-1}^2 + \sum z_{t-1}^2\right) + o_p(1)}.$$

When $T \to \infty$

$$\operatorname*{plim}_{T\to\infty} s^2 = \sigma_\xi^2 + \beta^2\frac{\sigma_\varepsilon^2\sigma_z^2}{\sigma_\varepsilon^2 + \sigma_z^2} \tag{B11}$$

Finally, we can write the $t-$statistics as

$$t_a = \hat{a}\left[s^2\left(\mathbf{X}'_{-1}\mathbf{X}_{-1}\right)^{-1}_{(1,1)}\right]^{-1/2},$$

$$t_b = \hat{b}\left[s^2\left(\mathbf{X}'_{-1}\mathbf{X}_{-1}\right)^{-1}_{(2,2)}\right]^{-1/2},$$

Following the same procedure as above, we find

$$\begin{aligned}
\operatorname*{plim}_{T\to\infty} t_a &= \operatorname*{plim}_{T\to\infty} \hat{a} \times \left(\operatorname*{plim}_{T\to\infty} s^2\right)^{-1/2} \times \operatorname*{plim}_{T\to\infty} \left(\left(\mathbf{X}'_{-1}\mathbf{X}_{-1}\right)^{-1}_{(11)}\right)^{-1/2} \\
&= \alpha\left(\sigma_\xi^2 + \beta^2\frac{\sigma_\varepsilon^2\sigma_z^2}{\sigma_\varepsilon^2 + \sigma_z^2}\right)^{-1/2}\left(\operatorname*{plim}_{T\to\infty}\frac{\sum x_{t-1}^2}{T\sum x_{t-1}^2 - \left(\sum x_{t-1}\right)^2}\right)^{-1/2} \\
&= \alpha\left(\sigma_\xi^2 + \beta^2\frac{\sigma_\varepsilon^2\sigma_z^2}{\sigma_\varepsilon^2 + \sigma_z^2}\right)^{-1/2}\left(\frac{\operatorname*{plim}_{T\to\infty}\left(\sum\varepsilon_{t-1}^2 + \sum z_{t-1}^2 + O_p(T^{1/2})\right)}{\operatorname*{plim}_{T\to\infty}\left(T\sum\varepsilon_{t-1}^2 + T\sum z_{t-1}^2 + O_p(T^{3/2})\right)}\right)^{-1/2} \\
&= \alpha\left(\sigma_\xi^2 + \beta^2\frac{\sigma_\varepsilon^2\sigma_z^2}{\sigma_\varepsilon^2 + \sigma_z^2}\right)^{-1/2}\left(\frac{\sigma_\varepsilon^2 + \sigma_z^2}{\operatorname*{plim}_{T\to\infty}\left(\sum\varepsilon_{t-1}^2 + \sum z_{t-1}^2 + O_p(T^{1/2})\right)}\right)^{-1/2} \\
\operatorname*{plim}_{T\to\infty} T^{-1/2}t_a &= \alpha\left(\sigma_\xi^2 + \beta^2\frac{\sigma_\varepsilon^2\sigma_z^2}{\sigma_\varepsilon^2 + \sigma_z^2}\right)^{-1/2}, && \text{(B12)}
\end{aligned}$$

and

$$
\begin{aligned}
\underset{T\to\infty}{\text{plim}}\ t_b &= \underset{T\to\infty}{\text{plim}}\ \hat{b} \times \left(\underset{T\to\infty}{\text{plim}}\ s^2\right)^{-1/2} \times \underset{T\to\infty}{\text{plim}}\ \left(\left(\mathbf{X}'_{-1}\mathbf{X}_{-1}\right)^{-1}_{(22)}\right)^{-1/2} \\
&= \left(\beta \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+\sigma_z^2}\right)\left(\sigma_\xi^2+\beta^2\frac{\sigma_\varepsilon^2\sigma_z^2}{\sigma_\varepsilon^2+\sigma_z^2}\right)^{-1/2}\left(\underset{T\to\infty}{\text{plim}}\ \frac{T}{T\sum x_{t-1}^2-\left(\sum x_{t-1}\right)^2}\right)^{-1/2} \\
&= \left(\beta \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+\sigma_z^2}\right)\left(\sigma_\xi^2+\beta^2\frac{\sigma_\varepsilon^2\sigma_z^2}{\sigma_\varepsilon^2+\sigma_z^2}\right)^{-1/2}\left(\frac{1}{\underset{T\to\infty}{\text{plim}}\ \left(\sum \varepsilon_{t-1}^2+\sum z_{t-1}^2+O_p(T^{1/2})\right)}\right)^{-1/2} \\
\underset{T\to\infty}{\text{plim}}\ T^{-1/2}t_b &= \left(\frac{\beta^2\sigma_\varepsilon^4}{\sigma_\xi^2\left(\sigma_\varepsilon^2+\sigma_z^2\right)+\beta^2\sigma_\varepsilon^2\sigma_z^2}\right)^{1/2}. \qquad \text{(B13)}
\end{aligned}
$$

B.2 If $\beta = 0$:

Note that the asymptotic behavior for $\hat{a}$ does not change since the terms with the largest order of divergence in (B3) and (B4) do not involve β. Similarly, the asymptotic behavior of $\hat{b}^{(d)}$ remains the same, yet the limit of $\hat{b}^{(n)}$ is different if $\beta = 0$. We find that

$$
\begin{aligned}
\hat{b} &= \frac{T\left(\sum \xi_t\varepsilon_{t-1}+\sum \xi_t z_{t-1}\right)+o_p(T^{3/2})}{T\left(\sum \varepsilon_{t-1}^2+\sum z_{t-1}^2\right)+O_p(T^{3/2})} \\
&= \frac{\frac{1}{T^{1/2}}\frac{1}{T^{1/2}}\left(\sum \xi_t\varepsilon_{t-1}+\sum \xi_t z_{t-1}\right)+o_p(T^{-1/2})}{\frac{1}{T}\left(\sum \varepsilon_{t-1}^2+\sum z_{t-1}^2\right)+O_p(T^{-1/2})} \\
T^{1/2}\hat{b} &= \frac{\frac{1}{T^{1/2}}\left(\sum \xi_t\left(\varepsilon_{t-1}+z_{t-1}\right)\right)+o_p(1)}{\frac{1}{T}\left(\sum \varepsilon_{t-1}^2+\sum z_{t-1}^2\right)+O_p(T^{-1/2})} \qquad \text{(B14)}
\end{aligned}
$$

The denominator of (B14) converges in probability to $\sigma_\varepsilon^2+\sigma_z^2$. The numerator involves the sum of the random variable $\xi_t\left(\varepsilon_{t-1}+z_{t-1}\right)$, which is an ergodic stationary m.d.s. with mean zero and constant variance $\sigma_\xi^2\left(\sigma_\varepsilon^2+\sigma_z^2\right)$. Thus, by the ESMD-CLT the numerator converges in distribution to $\mathcal{N}\left(0,\sigma_\xi^2\left(\sigma_\varepsilon^2+\sigma_z^2\right)\right)$. It follows that

$$
T^{1/2}\hat{b} \overset{D}{\to} \mathcal{N}\left(0,\frac{\sigma_\xi^2}{\sigma_\varepsilon^2+\sigma_z^2}\right). \qquad \text{(B15)}
$$

The asymptotic behavior of the numerator of s^2, i.e. $s^{2(n)}$ changes if $\beta = 0$, whereas for $s^{2(d)}$ there is no change. As a result, we get

$$
\underset{T\to\infty}{\text{plim}}\ s^2 = \underset{T\to\infty}{\text{plim}}\ \frac{\frac{1}{T^2}\left(\sum \xi_t^2\sum \varepsilon_{t-1}^2+\sum \xi_t^2\sum z_{t-1}^2\right)+o_p(1)}{\frac{1}{T}\left(\sum \varepsilon_{t-1}^2+\sum z_{t-1}^2\right)+o_p(1)} = \sigma_\xi^2.
$$

Plugging this result into the expression for t_a in (B12), it follows that $\operatorname*{plim}_{T\to\infty} T^{-1/2}t_a = \frac{\alpha}{\sigma_\xi}$. Finally, we find the asymptotic behavior of t_b in the case where $\beta = 0$. We re-write (B13) as

$$\begin{aligned} t_b &= \hat{b} \times \left(s^2 \left(\mathbf{X}'_{-1}\mathbf{X}_{-1}\right)^{-1}_{(22)} \right)^{-1/2} \\ &= \hat{b} \times \left(s^2 \frac{1}{\left(\sum \varepsilon^2_{t-1} + \sum z^2_{t-1} + O_p(T^{1/2})\right)} \right)^{-1/2} \\ &= T^{1/2}\hat{b} \times \left(s^2 \frac{1}{\left(\frac{1}{T}\sum \varepsilon^2_{t-1} + \sum z^2_{t-1} + O_p(T^{-1/2})\right)} \right)^{-1/2} \end{aligned} \quad \text{(B16)}$$

The first term in (B16), $T^{1/2}\hat{b}$ converges in distribution to a normal by (B15). The second term converges in probability to $\left(\frac{\sigma^2_\xi}{\sigma^2_\xi+\sigma^2_z}\right)^{-1/2}$. Hence,

$$t_b \xrightarrow{D} \mathcal{N}(0,1). \quad \text{(B17)}$$

C Proof of Theorem 2

This section presents proofs for the asymptotic results in Theorem 2. The IV estimator of regression model (12) is given by

$$\hat{\mathsf{b}}_{IV} \equiv \left(\hat{a},\ \hat{b}\right)' = \left(\mathbf{X}'_{-1}\mathbf{Q}_{-1}\left[\mathbf{Q}'_{-1}\mathbf{Q}_{-1}\right]^{-1}\mathbf{Q}'_{-1}\mathbf{X}_{-1}\right)^{-1}\left(\mathbf{X}'_{-1}\mathbf{Q}_{-1}\left[\mathbf{Q}'_{-1}\mathbf{Q}_{-1}\right]^{-1}\mathbf{Q}'_{-1}\mathbf{y}\right),$$

where $\mathbf{Q}_{-1}$, $\mathbf{X}'_{-1}$, and $\mathbf{y}$ are defined in Equations (20), (15), and (16), respectively. Introduce the following auxiliary notation

$$\mathbf{Q}'_{-1}\mathbf{X}_{-1} = \begin{pmatrix} T & \sum x_{t-1} \\ \sum q_{1,t-1} & \sum x_{t-1}q_{1,t-1} \\ \sum q_{2,t-1} & \sum x_{t-1}q_{2,t-1} \\ \vdots & \vdots \\ \sum q_{K,t-1} & \sum x_{t-1}q_{K,t-1} \end{pmatrix} \equiv \begin{pmatrix} \underset{1\times1}{T} & \underset{1\times1}{\chi} \\ \underset{K\times1}{\mathbf{q}} & \underset{K\times1}{\mathbf{r}} \end{pmatrix}$$

$$\mathbf{Q}'_{-1}\mathbf{y} = \begin{pmatrix} \sum y_t \\ \sum y_t q_{1,t-1} \\ \sum y_t q_{2,t-1} \\ \vdots \\ \sum y_t q_{K,t-1} \end{pmatrix} \equiv \begin{pmatrix} \underset{1\times1}{y} \\ \underset{K\times1}{\mathbf{t}} \end{pmatrix}$$

$$\mathbf{Q}'_{-1}\mathbf{Q}_{-1} = \begin{pmatrix} T & \sum q_{1,t-1} & \sum q_{2,t-1} & \cdots & \sum q_{K,t-1} \\ \sum q_{1,t-1} & \sum q^2_{1,t-1} & \sum q_{1,t-1}q_{2,t-1} & \cdots & \sum q_{1,t-1}q_{K,t-1} \\ \sum q_{2,t-1} & \sum q_{2,t-1}q_{1,t-1} & \sum q^2_{2,t-1} & \cdots & \sum q_{2,t-1}q_{K,t-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum q_{K,t-1} & \sum q_{K,t-1}q_{1,t-1} & \sum q_{K,t-1}q_{2,t-1} & \cdots & \sum q^2_{K,t-1} \end{pmatrix} \equiv \begin{pmatrix} \underset{1\times1}{T} & \underset{1\times K}{\mathbf{q}'} \\ \underset{K\times1}{\mathbf{q}} & \underset{K\times K}{\mathbf{B}} \end{pmatrix}$$

It follows that

$$(\mathbf{Q}_{-1}'\mathbf{Q}_{-1})^{-1} = \frac{1}{c}\begin{pmatrix} 1 & -\mathbf{q}'\mathbf{B}^{-1} \\ -\mathbf{B}^{-1}\mathbf{q} & c\mathbf{B}^{-1} + \mathbf{B}^{-1}\mathbf{q}\mathbf{q}'\mathbf{B}^{-1} \end{pmatrix},$$

where $\underset{1\times 1}{c} \equiv T - \mathbf{q}'\mathbf{B}^{-1}\mathbf{q}$. Furthermore,

$$\mathbf{X}_{-1}'\mathbf{Q}_{-1}(\mathbf{Q}_{-1}'\mathbf{Q}_{-1})^{-1} = \frac{1}{c}\begin{pmatrix} c & 0 \\ x - \mathbf{r}'\mathbf{B}^{-1}\mathbf{q} & -x\mathbf{q}'\mathbf{B}^{-1} + c\mathbf{r}'\mathbf{B}^{-1} + \mathbf{r}'\mathbf{B}^{-1}\mathbf{q}\mathbf{q}'\mathbf{B}^{-1} \end{pmatrix},$$

and

$$\begin{aligned}
\mathbf{X}_{-1}'\mathbf{Q}_{-1}(\mathbf{Q}_{-1}'\mathbf{Q}_{-1})^{-1}\mathbf{Q}_{-1}'\mathbf{X}_{-1} &= \frac{1}{c}\begin{pmatrix} Tc & xc \\ xc & \left(x - \mathbf{r}'\mathbf{B}^{-1}\mathbf{q}\right)\left(x - \mathbf{r}'\mathbf{B}^{-1}\mathbf{q}\right)' + c\mathbf{r}'\mathbf{B}^{-1}\mathbf{r} \end{pmatrix} \\
\mathbf{X}_{-1}'\mathbf{Q}_{-1}(\mathbf{Q}_{-1}'\mathbf{Q}_{-1})^{-1}\mathbf{Q}_{-1}'\mathbf{y} &= \frac{1}{c}\begin{pmatrix} yc \\ yx - y\mathbf{r}'\mathbf{B}^{-1}\mathbf{q} + \left(-x\mathbf{q}'\mathbf{B}^{-1} + c\mathbf{r}'\mathbf{B}^{-1} + \mathbf{r}'\mathbf{B}^{-1}\mathbf{q}\mathbf{q}'\mathbf{B}^{-1}\right)\mathbf{t} \end{pmatrix}.
\end{aligned}$$

Now, note that the following relation must hold true $\left(\mathbf{X}_{-1}'\mathbf{Q}_{-1}(\mathbf{Q}_{-1}'\mathbf{Q}_{-1})^{-1}\mathbf{Q}_{-1}'\mathbf{X}_{-1}\right)^{-1} =$

$$\frac{c}{Tc\left(x - \mathbf{r}'\mathbf{B}^{-1}\mathbf{q}\right)\left(x - \mathbf{r}'\mathbf{B}^{-1}\mathbf{q}\right)' + Tc^2\mathbf{r}'\mathbf{B}^{-1}\mathbf{r} - x^2c^2}\begin{pmatrix} \left(x - \mathbf{r}'\mathbf{B}^{-1}\mathbf{q}\right)\left(x - \mathbf{r}'\mathbf{B}^{-1}\mathbf{q}\right)' + c\mathbf{r}'\mathbf{B}^{-1}\mathbf{r} & -xc \\ -xc & Tc \end{pmatrix}. \tag{C1}$$

Thus, the IV estimate can be re-written as follows.

$$\hat{\mathsf{b}}_{IV} = \frac{1}{Tc\left(x - \mathbf{r}'\mathbf{B}^{-1}\mathbf{q}\right)\left(x - \mathbf{r}'\mathbf{B}^{-1}\mathbf{q}\right)' + Tc^2\mathbf{r}'\mathbf{B}^{-1}\mathbf{r} - x^2c^2} \times \tag{C2}$$

$$\begin{pmatrix} \left(x\mathbf{q}' - c\mathbf{r}' - \mathbf{r}'\mathbf{B}^{-1}\mathbf{q}\mathbf{q}'\right)c\mathbf{B}^{-1}\left(-y\mathbf{r} + x\mathbf{t}\right) \\ -\left(x\mathbf{q}' - c\mathbf{r}' - \mathbf{r}'\mathbf{B}^{-1}\mathbf{q}\mathbf{q}'\right)Tc\mathbf{B}^{-1}\mathbf{t} + \left(x\mathbf{q}' - T\mathbf{r}'\right)yc\mathbf{B}^{-1}\mathbf{q} \end{pmatrix} \tag{C3}$$

As for the proof of Theorem 1 in Section B, it is necessary to obtain the limit expression of the sums that appear in the definitions of the IV estimates and the associated t-ratios. Most of these expressions are summarized summarized in Table B1. The remaining sums can be found in Table C2.

From the expressions in Tables B1 and C2 it follows that all elements of $\mathbf{B}$ are of the order $O_p(T)$. Hence, it must hold that $\mathbf{B}^{-1}$ is of order $O_p(T^{-1})$. Furthermore, the elements of all vectors have the same convergence rates; i.e. $\mathbf{q}$ is $O_p(T^{1/2})$, $\mathbf{r}$ is $O_p(T)$, and $\mathbf{t}$ is $O_p(T)$ if $\beta \neq 0$ and $O_p(T^{1/2})$ otherwise. Finally, we note the following orders for the scalars. x is $O_p(T^{d+1/2})$, y is $O_p(T)$, and c is $O_p(T)$.

Let $\hat{a}^{(n)}$ denote the numerator of $\hat{a}$, given by (C3). Note that independently of the true value of β, we find the following dominant terms

$$\hat{a}^{(n)} = \underbrace{c^2y\mathbf{r}'\mathbf{B}^{-1}\mathbf{r}}_{O_p(T^4)} + o_p(T^4).$$

$\sum q_{k,t-1}$	$=$	$\rho_k \sum \varepsilon_{t-1} + \underbrace{\sum \upsilon_{k,t-1}}_{O_p(T^{1/2})}$
$\sum q^2_{k,t-1}$	$=$	$\underbrace{\rho_k^2 \sum \varepsilon^2_{t-1}}_{O_p(T)} + \underbrace{\sum \upsilon^2_{k,t-1}}_{O_p(T)} + 2\underbrace{\rho_k \sum \varepsilon_{t-1}\upsilon_{k,t-1}}_{O_p(T^{1/2})}$
$\sum y_t q_{k,t-1}$	$=$	$\alpha\rho_k \sum \varepsilon_{t-1} + \alpha \sum \upsilon_{k,t-1} + \beta\rho_k \sum \varepsilon^2_{t-1} + \beta \sum \varepsilon_{t-1}\upsilon_{k,t-1} + \rho_k \underbrace{\sum \xi_t \varepsilon_{t-1}}_{O_p(T^{1/2})} + \underbrace{\sum \xi_t \upsilon_{k,t-1}}_{O_p(T^{1/2})}$
$\sum x_{t-1} q_{k,t-1}$	$=$	$\rho_k \sum \varepsilon^2_{t-1} + \sum \varepsilon_{t-1}\upsilon_{k,t-1} + \rho_k \sum \varepsilon_{t-1} z_{t-1} + \underbrace{\sum z_{t-1}\upsilon_{k,t-1}}_{O_p(T^{1/2})}$
$\sum q_{k,t-1} q_{j,t-1}$	$=$	$\rho_k\rho_j \sum \varepsilon^2_{t-1} + \rho_k \sum \varepsilon_{t-1}\upsilon_{j,t-1} + \rho_j \sum \varepsilon_{t-1}\upsilon_{k,t-1} + \underbrace{\sum \upsilon_{k,t-1}\upsilon_{j,t-1}}_{O_p(T^{1/2})}$

Table C2: Expressions for sums in Theorem 2 with $j \neq k$; $k = 1, \cdots, K$.

Similarly, the denominator of $\hat{a}$ given in (C2) is

$$\hat{a}^{(d)} = \underbrace{T c^2 \mathbf{r}' \mathbf{B}^{-1} \mathbf{r}}_{O_p(T^4)} + o_p(T^4).$$

It follows that in the limit we can write

$$\underset{T\to\infty}{\text{plim}}\, \hat{a} = \underset{T\to\infty}{\text{plim}} \frac{\frac{1}{T^4} c^2 y \mathbf{r}' \mathbf{B}^{-1} \mathbf{r} + o_p(1)}{\frac{1}{T^3} c^2 \mathbf{r}' \mathbf{B}^{-1} \mathbf{r} + o_p(1)} = \underset{T\to\infty}{\text{plim}} \frac{y}{T} = \frac{\alpha T}{T} = \alpha \tag{C4}$$

Now, note that the denominator of $\hat{b}$ is identical to the denominator of $\hat{a}$. Hence $\hat{b}^{(d)} = \hat{a}^{(d)}$. For the limiting behavior of the numerator of $\hat{b}$, we need to distinguish between $\beta = 0$ and $\beta \neq 0$. First, define additional auxiliary variables. Let

$$\mathbf{p} \equiv \begin{pmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_K \end{pmatrix} \quad \mathbf{u} \equiv \begin{pmatrix} \sum \upsilon_{1,t-1} \\ \sum \upsilon_{2,t-1} \\ \vdots \\ \sum \upsilon_{K,t-1} \end{pmatrix} \quad \mathbf{v} \equiv \begin{pmatrix} \sum \upsilon^2_{1,t-1} \\ \sum \upsilon^2_{2,t-1} \\ \vdots \\ \sum \upsilon^2_{K,t-1} \end{pmatrix} \quad \mathbf{w} \equiv \begin{pmatrix} \sum \xi_t \upsilon_{1,t-1} \\ \sum \xi_t \upsilon_{2,t-1} \\ \vdots \\ \sum \xi_t \upsilon_{K,t-1} \end{pmatrix}.$$

C.1 Case 1 - If $\beta \neq 0$

Following the convergence rates in the tables, we find

$$\hat{b}^{(n)} = \underbrace{T c^2 \mathbf{r}' \mathbf{B}^{-1} \mathbf{t}}_{O_p(T^4)} + O_p(T^{7/2}).$$

Hence,

$$\underset{T\to\infty}{\text{plim}}\, \hat{b} = \underset{T\to\infty}{\text{plim}} \frac{\frac{1}{T^3} c^2 \mathbf{r}' \mathbf{B}^{-1} \mathbf{t} + O_p(T^{-1/2})}{\frac{1}{T^3} c^2 \mathbf{r}' \mathbf{B}^{-1} \mathbf{r} + o_p(1)} = \underset{T\to\infty}{\text{plim}} \frac{\mathbf{r}' \mathbf{B}^{-1} \mathbf{t}}{\mathbf{r}' \mathbf{B}^{-1} \mathbf{r}} = \underset{T\to\infty}{\text{plim}} \frac{\mathbf{r}' \mathbf{B}^{-1} \beta \sum \varepsilon^2_{t-1} \mathbf{p}}{\mathbf{r}' \mathbf{B}^{-1} \sum \varepsilon^2_{t-1} \mathbf{p}} = \beta \tag{C5}$$

Next we investigate the asymptotic behavior of s^2, which is defined as

$$s^2 = \frac{\left(\mathbf{y} - \mathbf{X}_{-1}\hat{\mathsf{b}}_{IV}\right)'\left(\mathbf{y} - \mathbf{X}_{-1}\hat{\mathsf{b}}_{IV}\right)}{T-2}.$$

Introduce the additional auxiliary notation $\hbar \equiv \sum y_t^2$, $o \equiv \sum x_{t-1}y_t$, and $m \equiv \sum x_t^2$. The respective orders are $O_p(T)$, $O_p(T)$ if $\beta \neq 0$ and $O_p(T^{d+1/2})$ otherwise, and $O_p(T)$. We can re-write s^2 as

$$s^2 = \frac{\hbar\hat{b}^{(d)} - 2y\hat{a}^{(n)} - 2o\hat{b}^{(n)} + T\hat{a}\hat{a}^{(n)} + 2\chi\hat{a}^{(n)}\hat{b} + m\hat{b}\hat{b}^{(n)}}{(T-2)\hat{b}^{(d)}}.$$

Note that the denominator of s^2, $s^{2(d)}$, is equal to $(T-2)a^{(d)} = (T-2)b^{(d)}$. For the case where $\beta \neq 0$, we can write the numerator of s^2, $s^{2(n)}$, in the limit as

$$s^{2(n)} = \underbrace{\hbar\hat{b}^{(d)} - 2y\hat{a}^{(n)} - 2o\hat{b}^{(n)} + T\hat{a}\hat{a}^{(n)} + m\hat{b}\hat{b}^{(n)}}_{O_p(T^5)} + o_p(T^5),$$

and hence by plugging in the dominant terms, we find

$$\begin{aligned}
\underset{T\to\infty}{\text{plim}}\, s^{2(n)} &= \underset{T\to\infty}{\text{plim}} \Big\{ \left(\alpha^2 T + \beta^2 \sum \varepsilon_{t-1}^2 + \sum \xi_t^2\right) Tc^2\mathbf{r}'\mathbf{B}^{-1}\mathbf{r} - 2\alpha Tc^2 y\mathbf{r}'\mathbf{B}^{-1}\mathbf{r} - 2\beta \sum \varepsilon_{t-1}^2 Tc^2\mathbf{r}'\mathbf{B}^{-1}\mathbf{t} \\
&\quad + T\alpha c^2 y\mathbf{r}'\mathbf{B}^{-1}\mathbf{r} + \left(\sum \varepsilon_{t-1}^2 + \sum z_{t-1}^2\right)\beta Tc^2\mathbf{r}'\mathbf{B}^{-1}\mathbf{t} + o_p(T^5)\Big\} \\
&= \underset{T\to\infty}{\text{plim}} \Big\{ \left(\alpha^2 T + \beta^2 \sum \varepsilon_{t-1}^2 + \sum \xi_t^2\right) T^3\mathbf{p}'\mathbf{B}^{-1}\mathbf{p}\left(\sum \varepsilon_{t-1}^2\right)^2 - 2\alpha^2 T^4\mathbf{p}'\mathbf{B}^{-1}\mathbf{p}\left(\sum \varepsilon_{t-1}^2\right)^2 \\
&\quad - 2\beta^2\left(\sum \varepsilon_{t-1}^2\right)^3 T^3\mathbf{p}'\mathbf{B}^{-1}\mathbf{p} + T^4\alpha^2\mathbf{p}'\mathbf{B}^{-1}\mathbf{p}\left(\sum \varepsilon_{t-1}^2\right)^2 \\
&\quad + \left(\sum \varepsilon_{t-1}^2 + \sum z_{t-1}^2\right)\beta^2 T^3\mathbf{p}'\mathbf{B}^{-1}\mathbf{p}\left(\sum \varepsilon_{t-1}^2\right)^2 + o_p(T^5)\Big\} \\
&= \underset{T\to\infty}{\text{plim}} \Big\{ T^3\mathbf{p}'\mathbf{B}^{-1}\mathbf{p}\left(\sum \varepsilon_{t-1}^2\right)^2\left(\sum \xi_t^2 + \beta^2\sum z_t^2\right) + o_p(T^5)\Big\}
\end{aligned}$$

We can therefore conclude that

$$\begin{aligned}
\underset{T\to\infty}{\text{plim}}\, s^2 &= \underset{T\to\infty}{\text{plim}} \frac{T^3\mathbf{p}'\mathbf{B}^{-1}\mathbf{p}\left(\sum \varepsilon_{t-1}^2\right)^2\left(\sum \xi_t^2 + \beta^2\sum z_t^2\right) + o_p(T^5)}{T^4\mathbf{p}'\mathbf{B}^{-1}\mathbf{p}\left(\sum \varepsilon_{t-1}^2\right)^2 + o_p(T^5)} \\
&= \underset{T\to\infty}{\text{plim}} \frac{T^{-2}\mathbf{p}'\mathbf{B}^{-1}\mathbf{p}\left(\sum \varepsilon_{t-1}^2\right)^2\left(\sum \xi_t^2 + \beta^2\sum z_t^2\right)}{T^{-1}\mathbf{p}'\mathbf{B}^{-1}\mathbf{p}\left(\sum \varepsilon_{t-1}^2\right)^2} \\
&= \underset{T\to\infty}{\text{plim}} \frac{1}{T}\left(\sum \xi_t^2 + \beta^2\sum z_t^2\right) = \sigma_\xi^2 + \beta^2\sigma_z^2. \qquad \text{(C6)}
\end{aligned}$$

Finally, we analyze the asymptotic behavior of the t-statistics. The expressions for $\left(\mathbf{X}'_{-1}\mathbf{X}_{-1}\right)^{-1}$ are given in Equation (C1).

$$\begin{aligned}
\operatorname*{plim}_{T\to\infty} t_a &= \operatorname*{plim}_{T\to\infty} \hat{a} \left(\operatorname*{plim}_{T\to\infty} s^2\right)^{-1/2} \left(\operatorname*{plim}_{T\to\infty} \left(\mathbf{X}'_{-1}\mathbf{Q}_{-1}\left[\mathbf{Q}'_{-1}\mathbf{Q}_{-1}\right]^{-1}\mathbf{Q}'_{-1}\mathbf{X}_{-1}\right)^{-1}_{(11)}\right)^{-1/2} \\
&= \alpha\left(\sigma_\xi^2+\beta^2\sigma_z^2\right)^{-1/2}\left(\operatorname*{plim}_{T\to\infty}\frac{\frac{1}{T}\hat{b}^{(d)}+\frac{1}{T}\varkappa^2 c}{\hat{b}^{(d)}}\right)^{-1/2} \\
&= \alpha\left(\sigma_\xi^2+\beta^2\sigma_z^2\right)^{-1/2}\left(\operatorname*{plim}_{T\to\infty}\frac{c^2\mathbf{r}'\mathbf{B}^{-1}\mathbf{r}+o_p(T^3)}{Tc^2\mathbf{r}'\mathbf{B}^{-1}\mathbf{r}+o_p(T^4)}\right)^{-1/2} \\
T^{-1/2}\operatorname*{plim}_{T\to\infty} t_a &= \alpha\left(\sigma_\xi^2+\beta^2\sigma_z^2\right)^{-1/2}\left(\operatorname*{plim}_{T\to\infty}\frac{T^{-3}c^2\mathbf{r}'\mathbf{B}^{-1}\mathbf{r}+o_p(1)}{T^{-3}c^2\mathbf{r}'\mathbf{B}^{-1}\mathbf{r}+o_p(1)}\right)^{-1/2} = \alpha\left(\sigma_\xi^2+\beta^2\sigma_z^2\right)^{-1/2}. \qquad \text{(C7)}
\end{aligned}$$

In a similar manner, we determine the asymptotics of t_b. To that end, we obtain an expression for $\mathbf{B}^{-1}$. We can write $\mathbf{B}=\sum\varepsilon_{t-1}^2\mathbf{p}\mathbf{p}'+\operatorname{diag}(\mathbf{v})+O_p(T^{1/2})$. Thus, we know that

$$\mathbf{B}^{-1}=(1/f)\left((\operatorname{diag}(\mathbf{v}))^{-1}f-(\operatorname{diag}(\mathbf{v}))^{-1}\sum\varepsilon_{t-1}^2\mathbf{p}\mathbf{p}'(\operatorname{diag}(\mathbf{v}))^{-1}\right)+O_p(T^{-1/2}),$$

where $f=1+\sum\varepsilon_{t-1}^2\mathbf{p}'(\operatorname{diag}(\mathbf{v}))^{-1}\mathbf{p}$. Then

$$\begin{aligned}
\operatorname*{plim}_{T\to\infty} t_b &= \operatorname*{plim}_{T\to\infty} \hat{b} \left(\operatorname*{plim}_{T\to\infty} s^2\right)^{-1/2} \left(\operatorname*{plim}_{T\to\infty} \left(\mathbf{X}'_{-1}\mathbf{Q}_{-1}\left[\mathbf{Q}'_{-1}\mathbf{Q}_{-1}\right]^{-1}\mathbf{Q}'_{-1}\mathbf{X}_{-1}\right)^{-1}_{(22)}\right)^{-1/2} \\
&= \beta\left(\sigma_\xi^2+\beta^2\sigma_z^2\right)^{-1/2}\left(\operatorname*{plim}_{T\to\infty}\frac{Tc^2}{\hat{b}^{(d)}}\right)^{-1/2} \\
&= \beta\left(\sigma_\xi^2+\beta^2\sigma_z^2\right)^{-1/2}\left(\operatorname*{plim}_{T\to\infty}\frac{T^3+O_p(T^2)}{Tc^2\mathbf{r}'\mathbf{B}^{-1}\mathbf{r}+o_p(T^4)}\right)^{-1/2} \\
&= \beta\left(\sigma_\xi^2+\beta^2\sigma_z^2\right)^{-1/2}\left(\operatorname*{plim}_{T\to\infty}\frac{T^{-1}}{T^{-1}\mathbf{p}'\mathbf{B}^{-1}\mathbf{p}\left(\sum\varepsilon_{t-1}^2\right)^2}\right)^{-1/2} \\
&= \beta\left(\sigma_\xi^2+\beta^2\sigma_z^2\right)^{-1/2}\left(\operatorname*{plim}_{T\to\infty}\mathbf{p}'\mathbf{B}^{-1}\mathbf{p}\left(\sum\varepsilon_{t-1}^2\right)^2\right)^{1/2} \\
&= \beta\left(\sigma_\xi^2+\beta^2\sigma_z^2\right)^{-1/2}\left(\operatorname*{plim}_{T\to\infty}\frac{\left(\sum\varepsilon_{t-1}^2\right)^2\mathbf{p}'(\operatorname{diag}(\mathbf{v}))^{-1}\mathbf{p}}{1+\sum\varepsilon_{t-1}^2\mathbf{p}'(\operatorname{diag}(\mathbf{v}))^{-1}\mathbf{p}}\right)^{1/2} \\
&= \beta\left(\sigma_\xi^2+\beta^2\sigma_z^2\right)^{-1/2}\left(\operatorname*{plim}_{T\to\infty}\frac{\left(\sum\varepsilon_{t-1}^2\right)^2\sum_{k=1}^K\rho_k^2\left(\sum\upsilon_{k,t-1}^2\right)^{-1}}{1+\sum\varepsilon_{t-1}^2\sum_{k=1}^K\rho_k^2\left(\sum\upsilon_{k,t-1}^2\right)^{-1}}\right)^{1/2} \\
T^{-1/2}\operatorname*{plim}_{T\to\infty} t_b &= \beta\left(\sigma_\xi^2+\beta^2\sigma_z^2\right)^{-1/2}\left(\frac{\sigma_\varepsilon^4\sum_{k=1}^K\frac{\rho_k^2}{\sigma_{\upsilon_k}^2}}{1+\sigma_\varepsilon^2\sum_{k=1}^K\frac{\rho_k^2}{\sigma_{\upsilon_k}^2}}\right)^{1/2} \qquad \text{(C8)}
\end{aligned}$$

C.2 Case 2 - If $\beta = 0$

If we let $\beta = 0$, the asymptotic behaviour of $\hat{a}$, $\hat{b}^{(d)}$, $s^{2(d)}$, $\underset{T\to\infty}{\text{plim}}\left(\mathbf{X}'_{-1}\mathbf{Q}_{-1}\left[\mathbf{Q}'_{-1}\mathbf{Q}_{-1}\right]^{-1}\mathbf{Q}'_{-1}\mathbf{X}_{-1}\right)^{-1}_{(11)}$, and $\underset{T\to\infty}{\text{plim}}\left(\mathbf{X}'_{-1}\mathbf{Q}_{-1}\left[\mathbf{Q}'_{-1}\mathbf{Q}_{-1}\right]^{-1}\mathbf{Q}'_{-1}\mathbf{X}_{-1}\right)^{-1}_{(22)}$ remains unaltered. However, the convergence of $\hat{b}^{(n)}$ and $s^{2(n)}$ changes. For the former, we find the following.

$$
\begin{aligned}
\hat{b}^{(n)} &= \underbrace{c\mathbf{r}'Tc\mathbf{B}^{-1}\mathbf{t} - T\mathbf{r}'yc\mathbf{B}^{-1}\mathbf{q}}_{O_p(T^{7/2})} + o_p(T^3) \\
&= \underbrace{Tc\mathbf{r}'\mathbf{B}^{-1}\left(T\mathbf{p}\sum \xi_t\varepsilon_{t-1} + T\mathbf{w}\right)}_{O_p(T^{7/2})} + o_p(T^3) \\
&= \underbrace{T^3\sum \varepsilon^2_{t-1}\mathbf{p}'\mathbf{B}^{-1}\left(\mathbf{p}\sum \xi_t\varepsilon_{t-1} + \mathbf{w}\right)}_{O_p(T^{7/2})} + O_p(T^3)
\end{aligned}
$$

Hence, in the limit we get

$$
\begin{aligned}
\underset{T\to\infty}{\text{plim}}\hat{b} &= \underset{T\to\infty}{\text{plim}}\frac{T^3\sum \varepsilon^2_{t-1}\mathbf{p}'\mathbf{B}^{-1}\left(\mathbf{p}\sum \xi_t\varepsilon_{t-1} + \mathbf{w}\right) + O_p(T^3)}{T^3\mathbf{p}'\sum \varepsilon^2_{t-1}\mathbf{B}^{-1}\sum \varepsilon^2_{t-1}\mathbf{p} + o_p(T^4)} \\
&= \underset{T\to\infty}{\text{plim}}\frac{T^{-1}\sum \varepsilon^2_{t-1}\mathbf{p}'\mathbf{B}^{-1}\left(\mathbf{p}\sum \xi_t\varepsilon_{t-1} + \mathbf{w}\right)}{T^{-1}\mathbf{p}'\sum \varepsilon^2_{t-1}\mathbf{B}^{-1}\sum \varepsilon^2_{t-1}\mathbf{p}} \\
&= \underset{T\to\infty}{\text{plim}}\frac{\mathbf{p}'\left(\text{diag}(\mathbf{v})\right)^{-1}\left(\mathbf{p}\sum \xi_t\varepsilon_{t-1} + \mathbf{w}\right)}{\mathbf{p}'\left(\text{diag}(\mathbf{v})\right)^{-1}\mathbf{p}\sum \varepsilon^2_{t-1}} \\
&= \underset{T\to\infty}{\text{plim}}\frac{\sum_{k=1}^K \rho_k^2\left(\sum \upsilon^2_{k,t-1}\right)^{-1}\left(\sum \xi_t\varepsilon_{t-1} + \frac{\sum \xi_t\upsilon_{k,t-1}}{\rho_k}\right)}{\sum_{k=1}^K \rho_k^2\left(\sum \upsilon^2_{k,t-1}\right)^{-1}\sum \varepsilon^2_{t-1}} \\
T^{1/2}\underset{T\to\infty}{\text{plim}}\hat{b} &= \underset{T\to\infty}{\text{plim}}\frac{\sum_{k=1}^K \rho_k^2\left(\frac{1}{T}\sum \upsilon^2_{k,t-1}\right)^{-1}\left(\frac{1}{T^{1/2}}\sum \xi_t\varepsilon_{t-1} + \frac{\frac{1}{T^{1/2}}\sum \xi_t\upsilon_{k,t-1}}{\rho_k}\right)}{\sum_{k=1}^K \rho_k^2\left(\frac{1}{T}\sum \upsilon^2_{k,t-1}\right)^{-1}\frac{1}{T}\sum \varepsilon^2_{t-1}} \\
&= \frac{\underset{T\to\infty}{\text{plim}}\frac{1}{T^{1/2}}\sum \xi_t\varepsilon_{t-1}}{\sigma^2_\varepsilon} + \frac{\sum_{k=1}^K \frac{\rho_k}{\sigma^2_{\upsilon_k}}\underset{T\to\infty}{\text{plim}}\frac{1}{T^{1/2}}\sum \xi_t\upsilon_{k,t-1}}{\sigma^2_\varepsilon\sum_{k=1}^K \frac{\rho_k^2}{\sigma^2_{\upsilon_k}}} \qquad \text{(C9)}
\end{aligned}
$$

Note that (C9) is the sum of two independent random variables that has zero mean and asymptotic variance given by

$$
\mathsf{Var}(T^{1/2}\underset{T\to\infty}{\text{plim}}\hat{b}) = \frac{\sigma^2_\xi\sigma^2_\varepsilon}{\sigma^4_\varepsilon} + \frac{\sum_{k=1}^K \frac{\rho_k^2}{\sigma^4_{\upsilon_k}}\sigma^2_\xi\sigma^2_{\upsilon_k}}{\sigma^4_\varepsilon\left(\sum_{k=1}^K \frac{\rho_k^2}{\sigma^2_{\upsilon_k}}\right)^2} = \frac{\sigma^2_\xi\left(\sigma^2_\varepsilon\sum_{k=1}^K \frac{\rho_k^2}{\sigma^2_{\upsilon_k}} + 1\right)}{\sigma^4_\varepsilon\sum_{k=1}^K \frac{\rho_k^2}{\sigma^2_{\upsilon_k}}}. \qquad \text{(C10)}
$$

Thus, by the ESMD-CLT the result in Theorem 2 follows suit. Now we consider the error variance s^2. For the case where $\beta = 0$, we can write the numerator of s^2, $s^{2(n)}$, in the limit as

$$\begin{aligned}
s^{2(n)} &= \underbrace{\hbar\hat{b}^{(d)} - 2y\hat{a}^{(n)} + T\hat{a}\hat{a}^{(n)}}_{O_p(T^5)} + o_p(T^{9/2}) \\
\operatorname*{plim}_{T\to\infty} s^{2(n)} &= \operatorname*{plim}_{T\to\infty} \left\{ \left(\alpha^2 T + \sum \xi_t^2 \right) T c^2 \mathbf{r}'\mathbf{B}^{-1}\mathbf{r} - 2\alpha T c^2 y \mathbf{r}'\mathbf{B}^{-1}\mathbf{r} + T\alpha c^2 y \mathbf{r}'\mathbf{B}^{-1}\mathbf{r} \right\} \\
&= \operatorname*{plim}_{T\to\infty} \left\{ \left(\alpha^2 T + \sum \xi_t^2 \right) T^3 \mathbf{p}'\mathbf{B}^{-1}\mathbf{p} \left(\sum \varepsilon_{t-1}^2 \right)^2 - 2\alpha^2 T^4 \mathbf{p}'\mathbf{B}^{-1}\mathbf{p} \left(\sum \varepsilon_{t-1}^2 \right)^2 \right. \\
&\quad \left. + T^4 \alpha^2 \mathbf{p}'\mathbf{B}^{-1}\mathbf{p} \left(\sum \varepsilon_{t-1}^2 \right)^2 \right\} \\
&= \operatorname*{plim}_{T\to\infty} \left\{ \sum \xi_t^2 T^3 \mathbf{p}'\mathbf{B}^{-1}\mathbf{p} \left(\sum \varepsilon_{t-1}^2 \right)^2 \right\}
\end{aligned}$$

Since the denominator is the same as in Case 1 in Section C.1 above, we can conclude that

$$\begin{aligned}
\operatorname*{plim}_{T\to\infty} s^2 &= \operatorname*{plim}_{T\to\infty} \frac{\sum \xi_t^2 T^3 \mathbf{p}'\mathbf{B}^{-1}\mathbf{p} \left(\sum \varepsilon_{t-1}^2 \right)^2 + o_p(T^{9/2})}{T^4 \mathbf{p}'\mathbf{B}^{-1}\mathbf{p} \left(\sum \varepsilon_{t-1}^2 \right)^2 + o_p(T^5)} \\
&= \operatorname*{plim}_{T\to\infty} \frac{T^{-2} \sum \xi_t^2 \mathbf{p}'\mathbf{B}^{-1}\mathbf{p} \left(\sum \varepsilon_{t-1}^2 \right)^2}{T^{-1} \mathbf{p}'\mathbf{B}^{-1}\mathbf{p} \left(\sum \varepsilon_{t-1}^2 \right)^2} = \sigma_\xi^2 \qquad \text{(C11)}
\end{aligned}$$

In the final step, we analyze the asymptotic behavior of the t-statistics. Note that the computation of the limiting behavior of t_a follows exactly the steps in (C7) in Section C.1 above, with $\operatorname*{plim}_{T\to\infty} s^2$ replaced by σ_ξ^2. Then

$$T^{-1/2} \operatorname*{plim}_{T\to\infty} t_a = \alpha \left(\sigma_\xi^2 \right)^{-1/2}. \qquad \text{(C12)}$$

For the derivation of t_b, we make the following replacements in Equation (C8) above. Replace $\operatorname*{plim}_{T\to\infty} s^2$ by σ_ξ^2 and $\operatorname*{plim}_{T\to\infty} \hat{b}$ by $T^{-1/2}$ times the expression in (C9). It follows that

$$\begin{aligned}
T^{-1/2} \operatorname*{plim}_{T\to\infty} t_b &= T^{-1/2} \frac{\sum_{k=1}^K \frac{\rho_k^2}{\sigma_{\upsilon_k}^2} \operatorname*{plim}_{T\to\infty} \left(\frac{1}{T^{1/2}} \sum \xi_t \varepsilon_{t-1} + \frac{\frac{1}{T^{1/2}} \sum \xi_t \upsilon_{k,t-1}}{\rho_k} \right)}{\sigma_\varepsilon^2 \sum_{k=1}^K \frac{\rho_k^2}{\sigma_{\upsilon_k}^2}} \left(\sigma_\xi^2 \right)^{-1/2} \left(\frac{\sigma_\varepsilon^4 \sum_{k=1}^K \frac{\rho_k^2}{\sigma_{\upsilon_k}^2}}{1 + \sigma_\varepsilon^2 \sum_{k=1}^K \frac{\rho_k^2}{\sigma_{\upsilon_k}^2}} \right)^{1/2} \\
\operatorname*{plim}_{T\to\infty} t_b &= \frac{\sum_{k=1}^K \frac{\rho_k^2}{\sigma_{\upsilon_k}^2} \operatorname*{plim}_{T\to\infty} \left(\frac{1}{T^{1/2}} \sum \xi_t \varepsilon_{t-1} + \frac{\frac{1}{T^{1/2}} \sum \xi_t \upsilon_{k,t-1}}{\rho_k} \right)}{\sigma_\varepsilon^2 \sum_{k=1}^K \frac{\rho_k^2}{\sigma_{\upsilon_k}^2}} \left(\sigma_\xi^2 \right)^{-1/2} \left(\frac{\sigma_\varepsilon^4 \sum_{k=1}^K \frac{\rho_k^2}{\sigma_{\upsilon_k}^2}}{1 + \sigma_\varepsilon^2 \sum_{k=1}^K \frac{\rho_k^2}{\sigma_{\upsilon_k}^2}} \right)^{1/2} \qquad \text{(C13)}
\end{aligned}$$

By the same logic as before, note that (C13) is a random variable with zero mean and asymptotic variance equal to

$$\mathsf{Var}\left\{\frac{\sum_{k=1}^{K}\frac{\rho_k^2}{\sigma_{\upsilon_k}^2}\operatorname*{plim}_{T\to\infty}\left(\frac{1}{T^{1/2}}\sum\xi_t\varepsilon_{t-1}+\frac{\frac{1}{T^{1/2}}\sum\xi_t\upsilon_{k,t-1}}{\rho_k}\right)}{\sigma_\varepsilon^2\sum_{k=1}^{K}\frac{\rho_k^2}{\sigma_{\upsilon_k}^2}}\right\}\frac{\sigma_\varepsilon^4\sum_{k=1}^{K}\frac{\rho_k^2}{\sigma_{\upsilon_k}^2}}{\sigma_\xi^2\left(1+\sigma_\varepsilon^2\sum_{k=1}^{K}\frac{\rho_k^2}{\sigma_{\upsilon_k}^2}\right)}$$

$$\frac{\sigma_\xi^2\left(1+\sigma_\varepsilon^2\sum_{k=1}^{K}\frac{\rho_k^2}{\sigma_{\upsilon_k}^2}\right)}{\sigma_\varepsilon^4\sum_{k=1}^{K}\frac{\rho_k^2}{\sigma_{\upsilon_k}^2}}\frac{\sigma_\varepsilon^4\sum_{k=1}^{K}\frac{\rho_k^2}{\sigma_{\upsilon_k}^2}}{\sigma_\xi^2\left(1+\sigma_\varepsilon^2\sum_{k=1}^{K}\frac{\rho_k^2}{\sigma_{\upsilon_k}^2}\right)}=1, \tag{C14}$$

where the last line follows directly from (C10).

D Proof of Corollary 1

This section presents proofs for the asymptotic results in Corollary 1. Sargan's $\mathcal{J}$ test for instrument validity is built upon a two-step procedure: (i) we estimate regression (12) by IV and, (ii) the resulting residuals, $\hat{\mathbf{e}}$, are in turn regressed on the instruments. The residuals of the second regression, $\hat{\mathbf{v}}$, as well as $\hat{\mathbf{e}}$ are then used to construct the $\mathcal{J}$ statistic, $\mathcal{J}=T\frac{\hat{\mathbf{e}}'\hat{\mathbf{e}}-\hat{\mathbf{v}}'\hat{\mathbf{v}}}{\hat{\mathbf{e}}'\hat{\mathbf{e}}}$, where $\hat{\mathbf{e}}=\mathbf{y}-\mathbf{X}_{-1}\hat{\mathfrak{b}}_{IV}$ and $\hat{\mathbf{v}}=\hat{\mathbf{e}}-\mathbf{Q}_{-1}\hat{\varpi}$. Note that the test statistic can be written as

$$\begin{aligned}\mathcal{J}&=T\frac{\hat{\mathbf{e}}'\hat{\mathbf{e}}-\hat{\mathbf{v}}'\hat{\mathbf{v}}}{\hat{\mathbf{e}}'\hat{\mathbf{e}}}\\&=\frac{\hat{\mathbf{e}}'\mathbf{Q}_{-1}(\mathbf{Q}_{-1}'\mathbf{Q}_{-1})^{-1}\mathbf{Q}_{-1}'\hat{\mathbf{e}}}{\frac{\hat{\mathbf{e}}'\hat{\mathbf{e}}}{T}}\\&=\frac{(-\beta\mathbf{z}+\boldsymbol{\xi})'\,\mathbf{Q}_{-1}\mathbf{L}}{\sqrt{\frac{\hat{\mathbf{e}}'\hat{\mathbf{e}}}{T}}}\left[\mathbf{I}-\mathbf{L}'\mathbf{Q}_{-1}'\mathbf{X}_{-1}\left(\mathbf{X}_{-1}'\mathbf{Q}_{-1}\mathbf{L}\mathbf{L}'\mathbf{Q}_{-1}'\mathbf{X}_{-1}\right)^{-1}\mathbf{X}_{-1}'\mathbf{Q}_{-1}\mathbf{L}\right]\frac{\mathbf{L}'\mathbf{Q}_{-1}'\left(-\beta\mathbf{z}+\boldsymbol{\xi}\right)}{\sqrt{\frac{\hat{\mathbf{e}}'\hat{\mathbf{e}}}{T}}},\end{aligned} \tag{D1}$$

where

$$\begin{aligned}\mathbf{z}'&\equiv\begin{pmatrix}z_1 & z_2 & z_3 & \dots & z_{T-1}\end{pmatrix}\\\boldsymbol{\xi}'&\equiv\begin{pmatrix}\xi_2 & \xi_3 & \xi_4 & \dots & \xi_T\end{pmatrix}\end{aligned}$$

and $\mathbf{L}$ is a $(K+1)\times(K+1)$ matrix such that $\mathbf{L}\mathbf{L}'=(\mathbf{Q}_{-1}'\mathbf{Q}_{-1})^{-1}$. We can write $\mathbf{L}$ as

$$\mathbf{L}=\begin{pmatrix}\frac{1}{\sqrt{c}} & 0\\-\frac{1}{\sqrt{c}}\mathbf{B}^{-1}\mathbf{q} & \mathbf{B}^{-1/2}\end{pmatrix}.$$

$\mathcal{J}$ in (D1) is the product of the transpose of a $(K+1)\times1$ vector multiplied by a $(K+1)\times(K+1)$ symmetric and idempotent matrix, $\left[\mathbf{I}-\mathbf{L}'\mathbf{Q}_{-1}'\mathbf{X}_{-1}\left(\mathbf{X}_{-1}'\mathbf{Q}_{-1}\mathbf{L}\mathbf{L}'\mathbf{Q}_{-1}'\mathbf{X}_{-1}\right)^{-1}\mathbf{X}_{-1}'\mathbf{Q}_{-1}\mathbf{L}\right]$, that has rank $K-1$, multiplied by the former $(K+1)\times1$ vector. If it can be proven that the $(K+1)\times1$ vector, $\frac{\mathbf{L}'\mathbf{Q}_{-1}'(-\beta\mathbf{z}+\boldsymbol{\xi})}{\sqrt{\frac{\hat{\mathbf{e}}'\hat{\mathbf{e}}}{T}}}$, converges to a standard normal distribution, the usual result follows and $\mathcal{J}\xrightarrow{D}\chi^2_{(K-1)}$.

We can express the vector $\frac{\mathbf{L}'\mathbf{Q}'_{-1}(-\beta\mathbf{z}+\boldsymbol{\xi})}{\sqrt{\frac{\hat{\mathbf{e}}'\hat{\mathbf{e}}}{T}}}$ as

$$\frac{\mathbf{L}'\mathbf{Q}'_{-1}\left(-\beta\mathbf{z}+\boldsymbol{\xi}\right)}{\sqrt{\frac{\hat{\mathbf{e}}'\hat{\mathbf{e}}}{T}}} = \frac{1}{\sqrt{\frac{\hat{\mathbf{e}}'\hat{\mathbf{e}}}{T}}}\begin{pmatrix} -\frac{\beta}{\sqrt{c}}\sum z_{t-1} + \frac{1}{\sqrt{c}}\sum \xi_t + \frac{\beta}{\sqrt{c}}\mathbf{q}'\mathbf{B}^{-1}\mathbf{s} - \frac{1}{\sqrt{c}}\mathbf{q}'\mathbf{B}^{-1}\mathbf{a} \\ -\beta\mathbf{B}^{-1/2}\mathbf{s} + \mathbf{B}^{-1/2}\mathbf{a} \end{pmatrix}, \tag{D2}$$

where

$$\mathbf{s} \equiv \begin{pmatrix} \sum z_{t-1}q_{1,t-1} \\ \sum z_{t-1}q_{2,t-1} \\ \vdots \\ \sum z_{t-1}q_{K,t-1} \end{pmatrix} \qquad \mathbf{a} \equiv \begin{pmatrix} \sum \xi_t q_{1,t-1} \\ \sum \xi_t q_{2,t-1} \\ \vdots \\ \sum \xi_t q_{K,t-1} \end{pmatrix}.$$

Both vectors, $\mathbf{s}$ and $\mathbf{a}$, are of the order $O_p(T^{1/2})$. Thus, if it were not for the first row, which is dominated by the $O_p(T^{d+1/2})$ sum $\sum z_{t-1}$, this vector would trivially converge to $\mathcal{N}(0,1)$[12]. Because of the first row of (D2), however, the ESMD-CLT does not apply and we cannot show the convergence result.

Now notice that the idempotent matrix that scales the vector in (D2) can be re-written as

$$\left[\mathbf{I} - \mathbf{L}'\mathbf{Q}'_{-1}\mathbf{X}_{-1}\left(\mathbf{X}'_{-1}\mathbf{Q}_{-1}\mathbf{L}\mathbf{L}'\mathbf{Q}'_{-1}\mathbf{X}_{-1}\right)^{-1}\mathbf{X}'_{-1}\mathbf{Q}_{-1}\mathbf{L}\right] =$$

$$\begin{pmatrix} \frac{\mathbf{q}'\mathbf{B}^{-1}\mathbf{q}}{T} - \frac{\mathbf{q}'\mathbf{B}^{-1}\left(\frac{\chi}{\sqrt{T}}\mathbf{q}-\sqrt{T}\mathbf{r}\right)\left(\frac{\chi}{\sqrt{T}}\mathbf{q}-\sqrt{T}\mathbf{r}\right)'\mathbf{B}^{-1}\mathbf{q}}{T(\chi-\mathbf{r}'\mathbf{B}^{-1}\mathbf{q})^2+Tc\mathbf{r}'\mathbf{B}^{-1}\mathbf{r}-\chi^2 c} & -\frac{\sqrt{c}}{T}\mathbf{q}'\mathbf{B}^{-1/2} + \frac{\sqrt{c}\mathbf{q}'\mathbf{B}^{-1}\left(\frac{\chi}{\sqrt{T}}\mathbf{q}-\sqrt{T}\mathbf{r}\right)\left(\frac{\chi}{\sqrt{T}}\mathbf{q}-\sqrt{T}\mathbf{r}\right)'\mathbf{B}^{-1/2}}{T(\chi-\mathbf{r}'\mathbf{B}^{-1}\mathbf{q})^2+Tc\mathbf{r}'\mathbf{B}^{-1}\mathbf{r}-\chi^2 c} \\ -\frac{\sqrt{c}}{T}\mathbf{B}^{-1/2}\mathbf{q} + \frac{\sqrt{c}\mathbf{B}^{-1/2}\left(\frac{\chi}{\sqrt{T}}\mathbf{q}-\sqrt{T}\mathbf{r}\right)\left(\frac{\chi}{\sqrt{T}}\mathbf{q}-\sqrt{T}\mathbf{r}\right)'\mathbf{B}^{-1}\mathbf{q}}{T(\chi-\mathbf{r}'\mathbf{B}^{-1}\mathbf{q})^2+Tc\mathbf{r}'\mathbf{B}^{-1}\mathbf{r}-\chi^2 c} & \frac{\mathbf{B}^{-1/2}\left(T\mathbf{B}-\mathbf{q}\mathbf{q}'\right)\mathbf{B}^{-1/2}}{T} - \frac{c\mathbf{B}^{-1/2}\left(\frac{\chi}{\sqrt{T}}\mathbf{q}-\sqrt{T}\mathbf{r}\right)\left(\frac{\chi}{\sqrt{T}}\mathbf{q}-\sqrt{T}\mathbf{r}\right)'\mathbf{B}^{-1/2}}{T(\chi-\mathbf{r}'\mathbf{B}^{-1}\mathbf{q})^2+Tc\mathbf{r}'\mathbf{B}^{-1}\mathbf{r}-\chi^2 c} \end{pmatrix}.$$

It follows that the in the limit we find the following.

$$\operatorname*{plim}_{T\to\infty}\left[\mathbf{I} - \mathbf{L}'\mathbf{Q}'_{-1}\mathbf{X}_{-1}\left(\mathbf{X}'_{-1}\mathbf{Q}_{-1}\mathbf{L}\mathbf{L}'\mathbf{Q}'_{-1}\mathbf{X}_{-1}\right)^{-1}\mathbf{X}'_{-1}\mathbf{Q}_{-1}\mathbf{L}\right] =$$

$$\begin{pmatrix} O_p(T^{-1})\left\{\frac{\operatorname*{plim}_{T\to\infty}\left(c\mathbf{q}'\mathbf{B}^{-1}\mathbf{q}\mathbf{r}'\mathbf{B}^{-1}\mathbf{r}-c(\mathbf{r}'\mathbf{B}^{-1}\mathbf{q})^2\right)}{\operatorname*{plim}_{T\to\infty}Tc\mathbf{r}'\mathbf{B}^{-1}\mathbf{r}+o_p(T^3)} \qquad \frac{\operatorname*{plim}_{T\to\infty}\left(T\sqrt{c}\mathbf{q}'\mathbf{B}^{-1}\mathbf{r}\mathbf{r}'\mathbf{B}^{-1/2}-T\sqrt{c}\mathbf{r}'\mathbf{B}^{-1}\mathbf{r}\mathbf{q}'\mathbf{B}^{-1/2}\right)+o_p(T^2)}{\operatorname*{plim}_{T\to\infty}Tc\mathbf{r}'\mathbf{B}^{-1}\mathbf{r}+o_p(T^3)}\right\}O_p(T^{-\frac{1}{2}}) \\ O_p(T^{-\frac{1}{2}})\left\{\frac{\operatorname*{plim}_{T\to\infty}\left(T\sqrt{c}\mathbf{B}^{-1/2}\mathbf{r}\mathbf{q}'\mathbf{B}^{-1}\mathbf{r}-T\sqrt{c}\mathbf{B}^{-1/2}\mathbf{q}\mathbf{r}'\mathbf{B}^{-1}\mathbf{r}\right)+o_p(T^2)}{\operatorname*{plim}_{T\to\infty}Tc\mathbf{r}'\mathbf{B}^{-1}\mathbf{r}+o_p(T^3)} \qquad \frac{\operatorname*{plim}_{T\to\infty}\mathbf{B}^{-1/2}\left(T\mathbf{B}c\mathbf{r}'\mathbf{B}^{-1}\mathbf{r}-Tc\mathbf{r}\mathbf{r}'\right)\mathbf{B}^{-1/2}+o_p(T^3)}{\operatorname*{plim}_{T\to\infty}Tc\mathbf{r}'\mathbf{B}^{-1}\mathbf{r}+o_p(T^3)}\right\}O_p(1) \end{pmatrix}$$

[12]From the results in Appendix C it follows that $\sqrt{\hat{\mathbf{e}}'\hat{\mathbf{e}}/T} \xrightarrow{P} \sqrt{\beta^2\sigma_z^2+\sigma_\xi^2}$.

This implies that the limit of the idempotent matrix and the $(K+1)\times 1$ vector is equal to

$$\begin{aligned}
&\underset{T\to\infty}{\text{plim}}\left\{\left[\mathbf{I}-\mathbf{L}'\mathbf{Q}_{-1}'\mathbf{X}_{-1}\left(\mathbf{X}_{-1}'\mathbf{Q}_{-1}\mathbf{L}\mathbf{L}'\mathbf{Q}_{-1}'\mathbf{X}_{-1}\right)^{-1}\mathbf{X}_{-1}'\mathbf{Q}_{-1}\mathbf{L}\right]\frac{\mathbf{L}'\mathbf{Q}_{-1}'\left(-\beta\mathbf{z}+\boldsymbol{\xi}\right)}{\sqrt{\frac{\hat{\mathbf{e}}'\hat{\mathbf{e}}}{T}}}\right\}\\
&=\underset{T\to\infty}{\text{plim}}\left[\mathbf{I}-\mathbf{L}'\mathbf{Q}_{-1}'\mathbf{X}_{-1}\left(\mathbf{X}_{-1}'\mathbf{Q}_{-1}\mathbf{L}\mathbf{L}'\mathbf{Q}_{-1}'\mathbf{X}_{-1}\right)^{-1}\mathbf{X}_{-1}'\mathbf{Q}_{-1}\mathbf{L}\right]\frac{\underset{T\to\infty}{\text{plim}}\mathbf{L}'\mathbf{Q}_{-1}'\left(-\beta\mathbf{z}+\boldsymbol{\xi}\right)}{\sqrt{\beta^2\sigma_z^2+\sigma_\xi^2}}\\
&=\frac{1}{\sqrt{\beta^2\sigma_z^2+\sigma_\xi^2}}\left(\begin{array}{l}\frac{\underset{T\to\infty}{\text{plim}}\left(T\sqrt{c}\mathbf{q}'\mathbf{B}^{-1}\mathbf{r}\mathbf{r}'\mathbf{B}^{-1/2}-T\sqrt{c}\mathbf{r}'\mathbf{B}^{-1}\mathbf{r}\mathbf{q}'\mathbf{B}^{-1/2}\right)}{\underset{T\to\infty}{\text{plim}}\,Tc\mathbf{r}'\mathbf{B}^{-1}\mathbf{r}}\underset{T\to\infty}{\text{plim}}\left(-\beta\mathbf{B}^{-1/2}\mathbf{s}+\mathbf{B}^{-1/2}\mathbf{a}\right)+o_p(T^{-\frac{1}{2}})\\ \frac{\underset{T\to\infty}{\text{plim}}\,\mathbf{B}^{-1/2}\left(T\mathbf{B}c\mathbf{r}'\mathbf{B}^{-1}\mathbf{r}-Tc\mathbf{r}\mathbf{r}'\right)\mathbf{B}^{-1/2}}{\underset{T\to\infty}{\text{plim}}\,Tc\mathbf{r}'\mathbf{B}^{-1}\mathbf{r}}\underset{T\to\infty}{\text{plim}}\left(-\beta\mathbf{B}^{-1/2}\mathbf{s}+\mathbf{B}^{-1/2}\mathbf{a}\right)+o_p(1)\end{array}\right)\\
&=\frac{1}{\sqrt{\beta^2\sigma_z^2+\sigma_\xi^2}}\left(\begin{array}{l}0\\ \\ \underset{T\to\infty}{\text{plim}}\left(I-\frac{\mathbf{B}^{-1/2}\mathbf{r}\mathbf{r}'\mathbf{B}^{-1}}{\mathbf{r}'\mathbf{B}^{-1}\mathbf{r}}\right)\underset{T\to\infty}{\text{plim}}\left(-\beta\mathbf{B}^{-1/2}\mathbf{s}+\mathbf{B}^{-1/2}\mathbf{a}\right)+o_p(1)\end{array}\right);
\end{aligned}$$

i.e. the first row of the vector $[\mathbf{L}'\mathbf{Q}_{-1}'(-\beta\mathbf{z}+\boldsymbol{\xi})]/\sqrt{\hat{\mathbf{e}}'\hat{\mathbf{e}}/T}$ in (D2) disappears in the limit. The remaining terms, $[\underset{T\to\infty}{\text{plim}}(-\beta\mathbf{B}^{-1/2}\mathbf{s}+\mathbf{B}^{-1/2}\mathbf{a})]/\sqrt{\beta^2\sigma_z^2+\sigma_\xi^2}$, are (scaled) sums of martingale difference sequences by Lemma 1. By the ESMD-CLT they converge to a standard normal distribution. Hence, $\mathcal{J}$ is asymptotically equivalent to

$$\begin{aligned}
&\left(\begin{array}{cc}0 & \mathbf{n}'\end{array}\right)\times\left(\begin{array}{ll}0 & 0\\ 0 & \underset{T\to\infty}{\text{plim}}\left(I-\frac{\mathbf{B}^{-1/2}\mathbf{r}\mathbf{r}'\mathbf{B}^{-1}}{\mathbf{r}'\mathbf{B}^{-1}\mathbf{r}}\right)\end{array}\right)\times\left(\begin{array}{c}0\\ \mathbf{n}\end{array}\right)\\
=\;&\mathbf{n}'\underset{T\to\infty}{\text{plim}}\left(I-\frac{\mathbf{B}^{-1/2}\mathbf{r}\mathbf{r}'\mathbf{B}^{-1}}{\mathbf{r}'\mathbf{B}^{-1}\mathbf{r}}\right)\mathbf{n}\sim\chi^2_{(K-1)},
\end{aligned}$$

where $\mathbf{n}$ is a $K\times 1$ vector that has a standard normal distribution, $\mathcal{N}(0,1)$. Since $\underset{T\to\infty}{\text{plim}}(I-\frac{\mathbf{B}^{-1/2}\mathbf{r}\mathbf{r}'\mathbf{B}^{-1}}{\mathbf{r}'\mathbf{B}^{-1}\mathbf{r}})$ is the probability limit of a $K\times K$ symmetric and idempotent matrix of rank $K-1$ the above result holds and it follows that

$$\mathcal{J}\overset{D}{\to}\chi^2_{(K-1)}. \tag{D3}$$

References

Andersen, T. B., Bollerslev, T., Diebold, F. X. and Ebens, H. (2001). The distribution of realized stock return volatility, *Journal of Financial Economics* **6**(1): 43–76.

Andersen, T. G. and Bollerslev, T. (1997). Heterogeneous information arrivals and return volatility dynamics: Uncovering the long-run in high frequency returns, *The Journal of Finance* **52**(3): 975–1005.

Andersen, T. G., Bollerslev, T. and Diebold, F. X. (2007). Roughing It Up: Including Jump Components in the Measurement, Modeling, and Forecasting of Return Volatility, *The Review of Economics and Statistics* **89**(4): 701–720.

Baillie, R. T., Bollerslev, T. and Mikkelsen, H. O. (1996). Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics* **74**(1): 3–30.

Bandi, F. M. and Perron, B. (2006). Long memory and the relation between implied and realized volatility, *Journal of Financial Econometrics* **4**(4): 636–670.

Banerjee, A., Dolado, J., Galbraith, J. W. and Hendry, D. F. (1993). *Co-Integration, Error-Correction, and the Econometric Analysis of Non-Stationary Data (Advanced Texts in Econometrics*, Oxford University Press.

Barndorff-Nielsen, O. E. and Shephard, N. (2002). Estimating Quadratic Variation Using Realised Variance, *Journal of Applied Econometrics* **17**: 457–477.

Barndorff-Nielsen, O. E. and Shephard, N. (2004). Power and Bipower Variation with Stochastic Volatility and Jumps, *Journal of Financial Econometrics* **2**(1): 1–37.

Beran, J. (1994). *Statistics for Long-Memory Processes*, Chapman and Hall, USA.

Beran, J., Feng, Y., Ghosh, S. and Kulik, R. (2013). *Long-Memory Processes*, Springer.

Billingsley, P. (1961). The Lindeberg-Levy Theorem for Martingales, *Proceedings of the American Mathematical Society*, Vol. 12, pp. 788–792.

Binsbergen, J. H. and Koijen, R. S. J. (2010). Predictive Regressions: A Present-Value Approach, *Journal of Finance* **65**(4): 1439–1471.

Bollerslev, T., Chou, R. Y. and Kroner, K. F. (1992). ARCH Modeling in Finance, *Journal of Econometrics* **52**: 5–59.

Bollerslev, T., Osterrieder, D., Sizova, N. and Tauchen, G. (2013). Risk and Return: Long-Run Relations, Fractional Cointegration, and Return Predictability, *Journal of Financial Economics* **108**: 409–424.

Bollerslev, T., Tauchen, G. and Sizova, N. (2012). Volatility in Equilibrium: Asymmetries and Dynamic Dependencies, *Review of Finance* **16**(73): 31–80. Available at SSRN: http://ssrn.com/abstract=1687985.

Bollerslev, T., Tauchen, G. and Zhou, H. (2009). Expected stock returns and variance risk premia, *Review of Financial Studies* **22**(11): 4463–4492.

Breitung, J. and Demetrescu, M. (2013). Instrumental Variable and Variable Addition Based Inference in Predictive Regressions, *Journal of Econometrics* **forthcoming**.

Campbell, J. Y. (2000). Asset pricing at the millennium, *The Journal of Finance* **55**(4): 1515–1567.

Campbell, J. Y. and Shiller, R. (1988a). The dividend-price ratio and expectations of future dividends and discount factors, *The Review of Financial Studies* **1**(3): 195–228.

Campbell, J. Y. and Shiller, R. (1988b). Stock prices, earnings, and expected dividends, *The Journal of Finance* **43**(3): 661–676.

Campbell, J. Y. and Vuolteenaho, T. (2004). Bad beta, good beta, *The American Economic Review* **94**(5): 1249–1275.

Campbell, J. Y. and Yogo, M. (2006). Efficient tests of stock return predictability, *Journal of Financial Economics* **81**: 27–60.

Carlini, F. and Santucci de Magistris, P. (2013). On the Identification of Fractionally Cointegrated VAR Models with the F(d) Condition, *CREATES Research Paper* **2013-44**.

Cavanagh, C., Elliott, G. and Stock, J. (1995). Inference in Models with Nearly Integrated Regressors, *Econometric Theory* **11**(5): 1131–1147.

Christensen, B. J. and Nielsen, M. O. (2006). Asymptotic normality of narrow-band least squares in the stationary fractional cointegration model and volatility forecasting, *Journal of Econometrics* **133**: 343–371.

Christensen, B. J. and Nielsen, M. O. (2007). The Effect of Long Memory in Volatility on Stock Market Fluctuations, *Review of Economics and Statistics* **89**(4): 684–700.

Cochrane, J. H. (1999). New facts in finance, *Economic Perspectives* **23**(3): 36–58.

Comte, F. and Renault, E. (1998). Long memory in continuous-time stochastic volatility models, *Mathematical Finance* **8**(4): 291–323.

Corsi, F., Pirino, D. and Reno, R. (2010). Threshold Bipower Variation and the Impact of Jumps on Volatility Forecasting, *Journal of Econometrics* **159**(2): 276–288.

Cuñado, J., Gil-Alana, L. A. and Perez de Garcia, F. (2005). A Test for Rational Bubbles in the NASDAQ Stock Index: A Fractionally Integrated Approach, *Journal of Banking and Finance* **29**(10): 2633–2654.

Deng, A. (2014). Understanding Spurious Regression in Financial Economics, *Journal of Financial Econometrics* **12**(1): 122–150.

Diebold, F. X. and Inoue, A. (2001). Long memory and regime switching, *Journal of Econometrics* **105**(1): 131–159.

Diebold, F. X. and Li, C. (2006). Forecasting the term structure of government bond yields, *Journal of Econometrics* **130**: 337–364.

Ding, Z., Granger, C. W. J. and Engle, R. F. (1993). A Long Memory Property of Stock Market Returns and a New Model, *Journal of Empirical Finance* **1**: 83–106.

Dolado, J. J., Gonzalo, J. and Mayoral, L. (2002). A Fractional Dickey-Fuller Test for Unit Roots, *Econometrica* **70**(5): 1963–2006.

Elliott, G. and Stock, J. H. (1994). Inference in Time Series Regression when the Order of Integration of a Regressor Is Unknown, *Econometric Theory* **10**: 672–700.

Engle, R. F. and Bollerslev, T. (1986). Modelling the Persistence of Conditional Variances, *Econometric Reviews* **5**: 1–50.

Engle, R. F., Lilien, D. M. and Robins, R. P. (1987). Estimating Time Varying Risk Premia in the Term Structure: the ARCH-M Model, *Econometrica* **55**: 391–407.

Fama, E. F. and French, K. R. (1988). Dividend yields and expected stock returns, *Journal of Financial Economics* **22**: 3–27.

Fama, E. F. and French, K. R. (1989). Business conditions and expected returns on stocks and bonds, *Journal of Financial Economics* **25**(1): 23–49.

Ferson, W. E., Sarkissian, S. and Simin, T. T. (2003). Spurious Regressions in Financial Economics?, *Journal of Finance* **58**(4): 1393–1414.

Geweke, J. and Porter-Hudak, S. (1983). The estimation and application of long memory time series models, *Journal of Time Series Analysis* **4**: 221–237.

Glosten, L. R., Jagannathan, R. and Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks, *The Journal of Finance* **48**(5): 1779–1801.

Gonzalo, J. and Pitarakis, J.-Y. (2012). Regime-Specific Predictability in Predictive Regressions, *Journal of Business and Economic Statistics* **30**(2): 229–241.

Granger, C. W. J. (1980). Long-Memory Relationships and the Aggregation of Dynamic Models, *Journal of Econometrics* **14**: 227–238.

Hamilton, J. D. (1994). *Time Series Analysis*, Princeton University Press, Princeton, New Jersey.

Henry, M. and Zaffaroni, P. (2002). The long range dependence paradigm for macroeconomics and finance, *Columbia University, Department of Economics, Discussion Paper Series* **0102-19**.

Huang, X. and Tauchen, G. (2005). The Relative Contribution of Jumps to Total Price Variance, *Journal of Financial Econometrics* **3**: 456–499.

Hurvich, C. M., Moulines, E. and Soulier, P. (2005). Estimating Long Memory in Volatility, *Econometrica* **73**(4): 1283–1328.

Jansson, M. and Moreira, M. (2006). Optimal Inference in Regression Models with Nearly Integrated Regressors, *Econometrica* **74**(3): 681–714.

Johansen, S. (2008). A representation theory for a class of vector autoregressive models for fractional processes, *Econometric Theory* **24**: 651676.

Johansen, S. (2009). Representation of cointegrated autoregressive processes with application to fractional processes, *Econometric Reviews* **28**: 121–145.

Johansen, S. and Nielsen, M. O. (2012). Likelihood Inference for a Fractionally Cointegrated Vector Autoregressive Model, *Econometrica* **80**(6): 2667–2732.

Kothari, S. P. and Shanken, J. (1997). Book-to-Market, Dividend Yield, and Expected Market Returns: A Time Series Analysis, *Journal of Financial Economics* **18**: 169–203.

Künsch, H. (1987). Statistical aspects of self-similar processes,, *in* Y. Prohorov and V. Sazonov (eds), *Proceedings of the First World Congress of the Bernoulli Society*, VNU Science Press, pp. 67–74.

Lettau, M. and Ludvigson, S. (2001). Consumption, aggregate wealth, and expected stock returns, *The Journal of Finance* **56**(3): 815–849.

Lettau, M. and van Nieuwerburgh, S. (2008). Reconciling the Return Predictabililty Evidence, *The Review of Financial Studies* **21**(4): 1607–1652.

Lewellen, J. (1999). The time-series relations among expected return, risk, and book-to-market, *Journal of Financial Economics* **54**: 5–43.

Lewellen, J. (2004). Predicting Returns with Financial Ratios, *Journal of Financial Economics* **74**: 209–235.

Lintner, J. (1965). The valuation of risky assets and the selection of risky investments in stock portfolios and capital budgets, *Review of Economics and Statistics* **47**: 13–37.

Magdalinos, T. and Phillips, P. C. B. (2009). Econometric Inference in the Vicinity of Unity, *Cowles Foundation Discussion Paper* **7**: Singapore Management University.

Markowitz, H. M. (1952). Portfolio selection, *Journal of Finance* **7**(1): 77–91.

Maynard, A. and Phillips, P. C. B. (2001). Rethinking an Old Empirical Puzzle: Econometric Evidence on the Forward Discount Anomaly, *Journal of Applied Econometrics* **16**: 671–708.

Maynard, A., Smallwood, A. and Wohar, M. E. (2013). Long Memory Regressors and Predictive Testing: A Two-Stage Rebalancing Approach, *Econometric Reviews* **32**(3): 318–360.

Meddahi, N. (2002). A Theoretical Comparison Between Integrated and Realized Volatility, *Journal of Applied Econometrics* **17**: 479–508.

Nielsen, M. O. and Morin, L. (2012). FCVARmodel.m: A Matlab Software Package for Estimation and Testing in the Fractionally Cointegrated VAR Model, *QED Working Paper* **1273**.

Nielsen, M. O. and Shimotsu, K. (2007). Determining the Cointegrating Rank in Nonstationary Fractional Systems by the Exact Local Whittle Approach, *Journal of Econometrics* **141**(2): 574–59.

Noriega, A. E. and Ventosa-Santaulària, D. (2007). Spurious Regression and Trending Variables, *Oxford Bulletin of Economics and Statistics* **69**(3): 439–444.

Pastor, L. and Stambaugh, R. F. (2009). Predictive Systems: Living with Imperfect Predictors, *Journal of Finance* **64**(4): 1583–1628.

Phillips, P. C. B. (1987). Towards a Unified Asymptotic Theory for Autoregression, *Biometrika* **74**(3): 535–547.

Phillips, P. C. B. (2012). On Confidence Intervals for Autoregressive Roots and Predictive Regression, *Cowles Foundation Discussion Paper* **1879**.

Phillips, P. C. B. and Lee, J. H. (2013). Predictive Regression under Various Degrees of Persistence and Robust Long-Horizon Regression, *Journal of Econometrics* **177**: 250–264.

Robinson, P. M. (1994). Efficient Tests of Nonstationary Hypotheses, *Journal of the American Statistical Association* **89**: 1420–1437.

Sargan, D. (1958). The Estimation of Economic Relationships Using Instrumental Variables, *Econometrica* **26**: 393–415.

Sharpe, W. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk, *Journal of Finance* **19**: 425–443.

Shimotsu, K. (2010). Exact Local Whittle Estimation of Fractional Integration with Unknown Mean and Time Trend, *Econometric Theory* **26**(2): 501–540.

Shimotsu, K. and Phillips, P. C. B. (2005). Exact local whittle estimation of fractional integration, *The Annals of Statistics* **33**: 1890–1933.

Stambaugh, R. F. (1986). Bias in Regressions with Lagged Stochastic Regressors, *University of Chicago, Graduate School of Business, Working Paper Series* **156**.

Stambaugh, R. F. (1999). Predictive Regressions, *Journal of Financial Economics* **54**(3): 375–421.

Stewart, C. (2011). A Note on Spurious Significance in Regressions Involving I(0) and I(1) Variables, *Empirical Economics* **41**(3): 565–571.

Stock, J. (1991). Confidence Intervals for the Largest Autoregressive Root in US Macroeconomic Time Series, *Journal of Monetary Economics* **28**(3): 435–459.

Sun, Y. and Phillips, P. C. B. (2004). Understanding the fisher equation, *Journal of Applied Econometrics* **19**(7): 869–886.

Tanaka, K. (1999). The Nonstationary Fractional Unit Root, *Econometric Theory* **15**: 549–582.

Tauchen, G. and Zhou, H. (2011). Realized Jumps on Financial Markets and Predicting Credit Spreads, *Journal of Econometrics* **160**: 102–118.

Torous, W. and Valkanov, R. (2000). Boundaries of Predictability: Noisy Preditive Regressions, *Working Paper - Anderson Graduate School of Management* .
URL: *http://escholarship.org/uc/item/33p7672z*

Tsay, W. J. and Chung, C. F. (2000). The Spurious Regression of Fractionally Integrated Processes, *Journal of Econometrics* **96**(1): 155–182.

Tables

Table 1: Size, Power, and (In-)Consistency of OLS Estimate $\hat{b}$

The table reports rejection rates in % at a nominal size of 5% based on a standard t-test of $H_0 : \beta = 0$ vs. $H_1 : \beta \neq 0$. We estimate Regression (12) by OLS. Simulations are based on 200,000 repetitions. The table also outlines the relative bias in the estimate, $(\hat{b}/\beta) \times 100$ in gray font. All errors are drawn from t-distributions.

		d															
		β															
		0.1				0.23				0.36				0.49			
$\sigma_\xi, \sigma_\varepsilon, \sigma_\eta$	T	-3	0	0.8	3	-3	0	0.8	3	-3	0	0.8	3	-3	0	0.8	3
1.7,1.7,1.7	250	99	5.1	94	99	99	5.1	92	99	99	5.1	83	98	90	5.1	48	90
		50		50	50	48		48	48	44		44	44	36		36	36
1.7,1.7,1.7	1000	100	5.0	100	100	100	5.0	100	100	100	5.1	99	100	99	5.0	94	99
		50		50	50	47		47	47	42		42	42	32		32	32
1.7,1.7,1.1	250	100	5.1	100	100	100	5.2	100	100	100	5.2	99	100	100	5.3	93	100
		68		68	68	66		66	66	62		62	62	54		54	54
1.7,1.7,1.1	1000	100	5.0	100	100	100	5.0	100	100	100	5.0	100	100	100	5.0	100	100
		69		69	69	66		66	66	61		61	61	50		50	50
1.7,1.1,1.7	250	98	5.1	64	98	98	5.1	57	98	95	5.1	35	95	67	5.2	9	67
		32		32	32	30		30	30	27		27	27	21		21	21
1.7,1.1,1.7	1000	100	5.1	99	100	100	5.1	98	100	100	5.0	95	100	97	5.0	48	97
		31		31	31	29		29	29	24		24	24	17		17	17
1.7,1.1,1.1	250	100	5.1	97	100	100	5.1	96	100	100	5.1	91	100	99	5.1	59	99
		50		50	50	48		48	48	43		43	43	36		36	36
1.7,1.1,1.1	1000	100	5.0	100	100	100	5.1	100	100	100	5.0	100	100	100	5.1	98	100
		50		50	50	47		47	47	41		41	41	31		31	31
1.1,1.7,1.7	250	99	5.2	98	99	99	5.1	98	99	99	5.1	95	99	92	5.1	72	92
		50		50	50	48		48	48	44		44	44	36		36	36
1.1,1.7,1.7	1000	100	4.9	100	100	100	5.0	100	100	100	5.0	100	100	99	5.1	98	99
		50		50	50	47		47	47	42		42	42	32		32	32
1.1,1.7,1.1	250	100	5.1	100	100	100	5.2	100	100	100	5.2	100	100	100	5.2	99	100
		68		68	68	66		66	66	62		62	62	54		54	54
1.1,1.7,1.1	1000	100	5.0	100	100	100	5.1	100	100	100	5.0	100	100	100	5.1	100	100
		69		69	69	66		66	66	61		61	61	50		50	50
1.1,1.1,1.7	250	99	5.1	90	99	98	5.2	86	98	96	5.1	69	96	73	5.2	26	73
		32		32	32	30		30	30	27		27	27	21		21	21
1.1,1.1,1.7	1000	100	5.0	100	100	100	5.1	99	100	100	5.1	99	100	98	5.1	80	98
		31		31	31	29		29	29	24		24	24	17		17	17
1.1,1.1,1.1	250	100	5.0	100	100	100	5.1	100	100	100	5.2	100	100	100	5.1	88	100
		50		50	50	48		48	48	43		43	43	36		36	36
1.1,1.1,1.1	1000	100	5.1	100	100	100	5.1	100	100	100	5.0	100	100	100	5.1	100	100
		50		50	50	47		47	47	41		41	41	31		31	31

Table 2: Size, Power, and Consistency of IV Estimate $\hat{b}$

The table reports rejection rates in % at a nominal size of 5% based on a standard t-test of $H_0 : \beta = 0$ vs. $H_1 : \beta \neq 0$. We estimate Regression (12) by IV. Simulations are based on 200,000 repetitions. The table also outlines the relative bias in the estimate, $(\hat{b}/\beta) \times 100$ in gray font. All errors are drawn from standard normal distributions.

		d / Corr(q_t, x_t^*) / β																	
		0.1						0.295						0.49					
		-0.8			0.65			-0.8			0.65			-0.8			0.65		
$\sigma_\xi, \sigma_\varepsilon, \sigma_\eta, \sigma_\upsilon$	T	-2	0	3	-2	0	3	-2	0	3	-2	0	3	-2	0	3	-2	0	3
1.7,1.7,1.7,1.7	250	100	4.7	100	100	4.4	100	100	4.7	100	100	4.3	100	100	4.5	100	100	4.0	100
		101		101	101		101	101		101	101		101	101		101	102		102
1.7,1.7,1.7,1.7	1000	100	4.9	100	100	4.9	100	100	4.9	100	100	4.8	100	100	4.9	100	100	4.7	100
		100		100	100		100	100		100	100		100	100		100	101		101
1.7,1.7,1.7,1.1	250	100	4.7	100	100	4.4	100	100	4.6	100	100	4.3	100	100	4.5	100	100	4.0	100
		101		101	101		101	101		101	101		101	101		101	102		102
1.7,1.7,1.7,1.1	1000	100	4.9	100	100	4.9	100	100	4.9	100	100	4.8	100	100	4.9	100	100	4.7	100
		100		100	100		100	100		100	100		100	100		100	101		101
1.7,1.7,1.1,1.7	250	100	4.9	100	100	4.6	100	100	4.8	100	100	4.7	100	100	4.8	100	100	4.5	100
		100		100	100		100	100		100	100		100	100		101	101		101
1.7,1.7,1.1,1.7	1000	100	5.0	100	100	4.9	100	100	5.0	100	100	4.9	100	100	5.0	100	100	4.8	100
		100		100	100		100	100		100	100		100	100		100	100		100
1.7,1.7,1.1,1.1	250	100	4.8	100	100	4.7	100	100	4.8	100	100	4.7	100	100	4.7	100	100	4.6	100
		100		100	100		100	100		100	100		100	101		101	101		101
1.7,1.7,1.1,1.1	1000	100	5.0	100	100	4.9	100	100	5.0	100	100	4.9	100	100	4.9	100	100	4.8	100
		100		100	100		100	100		100	100		100	100		100	100		100
1.7,1.1,1.7,1.7	250	100	4.3	100	100	3.8	100	100	4.2	100	100	3.6	100	100	3.8	100	99	3.0	99
		102		102	102		102	102		102	103		103	103		103	105		105
1.7,1.1,1.7,1.7	1000	100	4.8	100	100	4.7	100	100	4.8	100	100	4.6	100	100	4.7	100	100	4.4	100
		100		100	101		101	100		100	101		101	101		101	101		101
1.7,1.1,1.7,1.1	250	100	4.3	100	100	3.8	100	100	4.1	100	100	3.7	100	100	3.7	100	99	3.0	99
		102		102	102		102	102		102	103		103	103		103	105		105
1.7,1.1,1.7,1.1	1000	100	4.8	100	100	4.7	100	100	4.8	100	100	4.6	100	100	4.6	100	100	4.4	100
		100		100	101		101	100		100	101		101	101		101	101		101
1.7,1.1,1.1,1.7	250	100	4.7	100	100	4.4	100	100	4.6	100	100	4.4	100	100	4.4	100	100	4.1	100
		101		101	101		101	101		101	101		101	101		101	102		102
1.7,1.1,1.1,1.7	1000	100	4.9	100	100	4.8	100	100	4.9	100	100	4.8	100	100	4.8	100	100	4.7	100
		100		100	100		100	100		100	100		100	100		100	101		101
1.7,1.1,1.1,1.1	250	100	4.6	100	100	4.5	100	100	4.6	100	100	4.4	100	100	4.4	100	100	4.0	100
		101		101	101		101	101		101	101		101	101		101	102		102
1.7,1.1,1.1,1.1	1000	100	4.9	100	100	4.8	100	100	4.9	100	100	4.9	100	100	4.8	100	100	4.7	100
		100		100	100		100	100		100	100		100	100		100	101		101
1.1,1.7,1.7,1.7	250	100	4.7	100	100	4.4	100	100	4.6	100	100	4.3	100	100	4.5	100	100	4.0	100
		101		101	101		101	101		101	101		101	101		101	102		102
1.1,1.7,1.7,1.7	1000	100	4.9	100	100	4.9	100	100	4.9	100	100	4.8	100	100	4.9	100	100	4.7	100
		100		100	100		100	100		100	100		100	100		100	101		101
1.1,1.7,1.7,1.1	250	100	4.7	100	100	4.4	100	100	4.6	100	100	4.4	100	100	4.4	100	100	4.1	100
		101		101	101		101	101		101	101		101	101		101	102		102
1.1,1.7,1.7,1.1	1000	100	4.9	100	100	4.8	100	100	4.9	100	100	4.8	100	100	4.8	100	100	4.7	100
		100		100	100		100	100		100	100		100	100		100	101		101
1.1,1.7,1.1,1.7	250	100	4.8	100	100	4.7	100	100	4.8	100	100	4.7	100	100	4.7	100	100	4.6	100
		100		100	100		100	100		100	100		100	101		101	101		101
1.1,1.7,1.1,1.7	1000	100	5.0	100	100	4.9	100	100	5.0	100	100	4.9	100	100	4.9	100	100	4.8	100
		100		100	100		100	100		100	100		100	100		100	100		100
1.1,1.7,1.1,1.1	250	100	4.8	100	100	4.7	100	100	4.8	100	100	4.7	100	100	4.7	100	100	4.5	100
		100		100	100		100	100		100	100		100	101		101	101		101
1.1,1.7,1.1,1.1	1000	100	5.0	100	100	4.9	100	100	4.9	100	100	5.0	100	100	4.9	100	100	4.9	100
		100		100	100		100	100		100	100		100	100		100	100		100
1.1,1.1,1.7,1.7	250	100	4.3	100	100	3.8	100	100	4.1	100	100	3.7	100	100	3.7	100	99	3.0	99
		102		102	102		102	102		102	103		103	103		103	105		105

Table 2 – continued from previous page

1.1,1.1,1.7,1.7	1000	100	4.8	100	100	4.7	100	100	4.8	100	100	4.6	100	100	4.6	100	100	4.4	100
		100		100	101		101	100		100	101		101	101		101	101		101
1.1,1.1,1.7,1.1	250	100	4.2	100	100	3.8	100	100	4.2	100	100	3.7	100	100	3.7	100	99	3.0	99
		102		102	102		102	102		102	103		103	103		103	105		105
1.1,1.1,1.7,1.1	1000	100	4.8	100	100	4.7	100	100	4.8	100	100	4.7	100	100	4.6	100	100	4.4	100
		100		100	101		101	100		100	101		101	101		101	101		101
1.1,1.1,1.1,1.7	250	100	4.6	100	100	4.5	100	100	4.6	100	100	4.4	100	100	4.4	100	100	4.0	100
		101		101	101		101	101		101	101		101	101		101	102		102
1.1,1.1,1.1,1.7	1000	100	4.9	100	100	4.8	100	100	4.9	100	100	4.9	100	100	4.8	100	100	4.7	100
		100		100	100		100	100		100	100		100	100		100	101		101
1.1,1.1,1.1,1.1	250	100	4.7	100	100	4.5	100	100	4.6	100	100	4.3	100	100	4.4	100	100	4.1	100
		101		101	101		101	101		101	101		101	101		101	102		102
1.1,1.1,1.1,1.1	1000	100	4.9	100	100	4.9	100	100	4.9	100	100	4.8	100	100	4.8	100	100	4.7	100
		100		100	100		100	100		100	100		100	100		100	101		101

Table 3: Size, Power, and (In-)Consistency of IV Estimate $\hat{b}$ with Irrelevant Instrument

The table reports rejection rates in % at a nominal size of 5% based on a standard t-test of $H_0 : \beta = 0$ vs. $H_1 : \beta \neq 0$. We estimate Regression (12) by IV using an irrelevant instrument. That is, we set $\mathsf{Corr}(q_t, x_t^*)$ equal to zero. Simulations are based on 200,000 repetitions. The table also outlines the relative bias in the estimate, $(\hat{b}/\beta) \times 100$ in gray font. All errors are drawn from continuous uniform distributions.

		d								
		β								
		0.1			0.295			0.49		
$\sigma_\xi, \sigma_\varepsilon, \sigma_\eta, \sigma_\upsilon$	T	-2	0	3	-2	0	3	-2	0	3
1.7,1.7,1.7,1.7	250	4	0.01	5	3	0.01	4	2	0.01	3
		-218		-135	73		25	86		80
1.7,1.7,1.7,1.7	1000	4	0.01	5	3	0.01	4	2	0.01	2
		23		127	33		58	-84		6
1.7,1.7,1.7,1.1	250	4	0.01	5	3	0.02	5	2	0.01	3
		-122		-130	71		-1522	157		133
1.7,1.7,1.7,1.1	1000	4	0.01	5	3	0.01	4	2	0.01	2
		-526		278	25		46	-105		-80
1.7,1.7,1.1,1.7	250	9	0.01	13	8	0.01	11	6	0.01	7
		-412		1472	53		71	434		81
1.7,1.7,1.1,1.7	1000	9	0.01	13	8	0.01	11	4	0.01	6
		76		-414	-10		-228	412		-91
1.7,1.7,1.1,1.1	250	9	0.01	13	8	0.01	11	6	0.01	7
		-539		429	74		-27	121		406
1.7,1.7,1.1,1.1	1000	9	0.01	12	8	0.01	11	4	0.01	6
		-335		86	-56		294	483		54
1.7,1.1,1.7,1.7	250	1	0.01	1	1	0.01	1	0	0.01	1
		-13		18	358		-101	162		76
1.7,1.1,1.7,1.7	1000	1	0.01	1	1	0.01	1	0	0.01	0
		-294		65	-2227		-2268	147		7
1.7,1.1,1.7,1.1	250	1	0.01	1	1	0.01	1	0	0.01	1
		-2		-13	-73		-77	81		-10
1.7,1.1,1.7,1.1	1000	1	0.01	1	1	0.01	1	0	0.01	0
		-333		108	-2316		-2307	57		11
1.7,1.1,1.1,1.7	250	2	0.01	4	2	0.01	3	1	0.01	2
		-70		95	-66		-1928	247		71
1.7,1.1,1.1,1.7	1000	2	0.01	4	2	0.01	3	1	0.01	2
		148		59	68		23	-147		103

Table 3 – continued from previous page

1.7,1.1,1.1,1.1	250	2	0.01	4	2	0.01	3	1	0.01	2
		140		426	-134		1974	27		-5
1.7,1.1,1.1,1.1	1000	2	0.01	4	2	0.01	3	1	0.01	1
		17		56	-10		-142	28		91
1.1,1.7,1.7,1.7	250	5	0.01	6	4	0.02	5	3	0.01	3
		-125		-124	61		-1367	141		145
1.1,1.7,1.7,1.7	1000	5	0.01	6	4	0.01	5	2	0.01	2
		-391		265	32		48	-82		-87
1.1,1.7,1.7,1.1	250	5	0.01	6	5	0.01	5	3	0.01	3
		-94		106	-633		-1419	201		61
1.1,1.7,1.7,1.1	1000	5	0.01	6	4	0.01	5	2	0.01	2
		200		49	59		16	-119		86
1.1,1.7,1.1,1.7	250	13	0.01	15	11	0.01	13	7	0.01	9
		-122		251	74		-8	113		408
1.1,1.7,1.1,1.7	1000	13	0.01	15	11	0.01	12	6	0.01	7
		-352		95	-92		262	364		48
1.1,1.7,1.1,1.1	250	13	0.01	15	11	0.01	13	7	0.01	9
		-593		-179	80		81	415		16
1.1,1.7,1.1,1.1	1000	13	0.01	15	11	0.01	12	6	0.01	7
		139		146	117		446	17		-278
1.1,1.1,1.7,1.7	250	1	0.01	2	1	0.01	1	1	0.01	1
		2		-14	-79		-70	80		-14
1.1,1.1,1.7,1.7	1000	1	0.01	2	1	0.01	1	0	0.01	1
		-250		66	-2306		-2332	47		18
1.1,1.1,1.7,1.1	250	1	0.01	2	1	0.01	1	1	0.01	1
		-17		-5	-37		-470	-33		-36
1.1,1.1,1.7,1.1	1000	1	0.01	2	1	0.01	1	0	0.01	1
		-133		99	-2452		-9	47		120
1.1,1.1,1.1,1.7	250	4	0.01	5	3	0.01	4	2	0.01	3
		131		351	-506		1839	36		-11
1.1,1.1,1.1,1.7	1000	4	0.01	5	3	0.01	4	2	0.01	2
		26		51	-3		-134	43		85
1.1,1.1,1.1,1.1	250	4	0.01	5	3	0.01	4	2	0.01	3
		-3		426	1202		26	-38		27
1.1,1.1,1.1,1.1	1000	4	0.01	5	3	0.01	4	2	0.01	2
		27		-327	-93		53	58		-11

Table 4: Size, Power, and (In-)Consistency of IV Estimate $\hat{b}$ with Invalid Instruments

The table reports rejection rates in % at a nominal size of 5% based on a standard t-test of $H_0 : \beta = 0$ vs. $H_1 : \beta \neq 0$. We estimate Regression (12) by IV using an invalid instrument of type 1 and type 2. That is, we set $\mathsf{Corr}(q_{t-1}, z_{t-1}) = 0.7$ for the invalid instrument of type 1; for the invalid instrument of type 2, we let $\mathsf{Corr}(q_{t-1}, \xi_t) = 0.7$. Both instruments are relevant with $\mathsf{Corr}(q_t, x_t^*) = 0.7$. Simulations are based on 200,000 repetitions. The table also outlines the relative bias in the estimate, $(\hat{b}/\beta) \times 100$ in gray font. All errors are drawn from standard normal distributions.

		d																	
		Invalid Instrument																	
		β																	
		0.1						0.295						0.49					
		type 1			type 2			type 1			type 2			type 1			type 2		
$\sigma_\xi, \sigma_\varepsilon, \sigma_\eta, \sigma_\mu$	T	-2	0	3	-2	0	3	-2	0	3	-2	0	3	-2	0	3	-2	0	3
1.7,1.7,1.7,1.7	250	100	5.1	100	100	100	87	100	5.1	100	100	100	79	100	5.0	100	99	100	51
		50		50	50		135	49		49	50		135	69		69	50		136
1.7,1.7,1.7,1.7	1000	100	5.0	100	100	100	100	100	5.0	100	100	100	100	100	5.0	100	100	100	99

		50		50	50		134	48		48	50		134	65		65	50		134
1.7,1.7,1.7,1.1	250	100	5.1	100	100	100	87	100	5.1	100	100	100	79	100	5.0	100	99	100	51
		50		50	50		135	49		49	50		135	69		69	50		136
1.7,1.7,1.7,1.1	1000	100	5.0	100	100	100	100	100	5.0	100	100	100	100	100	5.0	100	100	100	99
		50		50	50		134	48		48	50		134	65		65	50		134
1.7,1.7,1.1,1.7	250	100	5.1	100	100	100	100	100	5.1	100	100	100	99	99	5.1	99	100	100	93
		60		60	50		134	60		60	50		134	77		77	50		135
1.7,1.7,1.1,1.7	1000	100	5.0	100	100	100	100	100	5.0	100	100	100	100	100	5.0	100	100	100	100
		60		60	50		134	59		59	50		134	74		74	50		134
1.7,1.7,1.1,1.1	250	100	5.1	100	100	100	100	100	5.1	100	100	100	99	99	5.1	99	100	100	94
		60		60	50		134	60		60	50		134	77		77	50		135
1.7,1.7,1.1,1.1	1000	100	5.0	100	100	100	100	100	5.0	100	100	100	100	100	5.0	100	100	100	100
		60		60	50		134	59		59	50		134	74		74	50		134
1.7,1.1,1.7,1.7	250	100	5.1	100	100	100	80	100	5.1	100	100	100	68	100	4.9	100	99	100	22
		40		40	23		155	39		39	24		155	60		60	24		158
1.7,1.1,1.7,1.7	1000	100	5.0	100	100	100	100	100	5.0	100	100	100	100	100	5.0	100	100	100	99
		39		39	24		152	38		38	24		152	55		55	24		153
1.7,1.1,1.7,1.1	250	100	5.1	100	100	100	80	100	5.1	100	100	100	68	100	4.9	100	99	100	22
		40		40	23		155	39		39	24		155	60		60	24		158
1.7,1.1,1.7,1.1	1000	100	5.0	100	100	100	100	100	5.0	100	100	100	100	100	5.0	100	100	100	99
		39		39	24		152	38		38	24		152	55		55	24		153
1.7,1.1,1.1,1.7	250	100	5.1	100	100	100	100	100	5.1	100	100	100	99	99	5.0	100	100	100	90
		50		50	23		153	49		49	23		153	69		69	23		154
1.7,1.1,1.1,1.7	1000	100	5.0	100	100	100	100	100	5.1	100	100	100	100	100	5.0	100	100	100	100
		50		50	23		151	48		48	24		151	65		65	24		152
1.7,1.1,1.1,1.1	250	100	5.1	100	100	100	100	100	5.2	100	100	100	99	99	5.0	100	100	100	90
		50		50	23		153	49		49	23		153	69		69	23		154
1.7,1.1,1.1,1.1	1000	100	5.0	100	100	100	100	100	5.0	100	100	100	100	100	5.1	100	100	100	100
		50		50	23		151	48		48	24		151	65		65	24		152
1.1,1.7,1.7,1.7	250	100	5.1	100	96	100	46	100	5.1	100	94	100	37	100	5.0	100	85	100	17
		50		50	68		123	49		49	68		123	69		69	68		124
1.1,1.7,1.7,1.7	1000	100	5.0	100	100	100	99	100	5.0	100	100	100	98	100	5.0	100	100	100	83
		50		50	67		122	48		48	67		122	65		65	68		122
1.1,1.7,1.7,1.1	250	100	5.1	100	96	100	46	100	5.1	100	94	100	37	100	5.0	100	85	100	17
		50		50	68		123	49		49	68		123	69		69	68		124
1.1,1.7,1.7,1.1	1000	100	5.0	100	100	100	99	100	5.1	100	100	100	98	100	5.0	100	100	100	83
		50		50	67		122	48		48	67		122	65		65	68		122
1.1,1.7,1.1,1.7	250	100	5.1	100	100	100	90	100	5.1	100	100	100	84	99	5.1	100	98	100	62
		60		60	67		122	60		60	67		123	77		77	67		123
1.1,1.7,1.1,1.7	1000	100	5.0	100	100	100	100	100	5.0	100	100	100	100	100	5.0	100	100	100	99
		60		60	67		122	59		59	67		122	74		74	67		122
1.1,1.7,1.1,1.1	250	100	5.1	100	100	100	90	100	5.1	100	100	100	84	100	5.1	100	98	100	62
		60		60	67		122	60		60	67		123	77		77	67		123
1.1,1.7,1.1,1.1	1000	100	5.0	100	100	100	100	100	5.0	100	100	100	100	100	5.1	100	100	100	99
		60		60	67		122	59		59	67		122	74		74	67		122
1.1,1.1,1.7,1.7	250	100	5.1	100	97	100	30	100	5.1	100	96	100	18	100	4.9	100	89	100	2
		40		40	51		136	39		39	51		137	60		60	52		139
1.1,1.1,1.7,1.7	1000	100	5.0	100	100	100	99	100	5.0	100	100	100	98	100	5.0	100	100	100	78
		39		39	50		134	38		38	50		134	55		55	50		135
1.1,1.1,1.7,1.1	250	100	5.1	100	97	100	30	100	5.1	100	96	100	18	100	4.9	100	89	100	2
		40		40	51		137	39		39	51		137	60		60	52		139
1.1,1.1,1.7,1.1	1000	100	5.0	100	100	100	99	100	5.0	100	100	100	98	100	5.1	100	100	100	78
		39		39	50		134	38		38	50		134	55		55	50		135
1.1,1.1,1.1,1.7	250	100	5.1	100	100	100	87	100	5.2	100	100	100	79	100	5.0	100	99	100	51
		50		50	50		135	49		49	50		135	69		69	50		136
1.1,1.1,1.1,1.7	1000	100	5.0	100	100	100	100	100	5.0	100	100	100	100	100	5.1	100	100	100	99
		50		50	50		134	48		48	50		134	65		65	50		134
1.1,1.1,1.1,1.1	250	100	5.1	100	100	100	87	100	5.1	100	100	100	79	100	4.9	100	99	100	51
		50		50	50		135	49		49	50		135	69		69	50		136
1.1,1.1,1.1,1.1	1000	100	5.0	100	100	100	100	100	5.1	100	100	100	100	100	5.0	100	100	100	99
		50		50	50		134	48		48	50		134	65		65	50		134

Table 5: Size and Power of $\mathcal{J}-$test

The table reports rejection rates in % at a nominal size of 5% based on a $\mathcal{J}$-test of H_0 : valid instruments vs. H_1 : invalid instruments of type 2. We estimate Regression (12) by IV with $K = 2$ and subsequently estimate Regression (24) by OLS to compute the $\mathcal{J}$-statistic. $q_{1,t}$ and $q_{2,t}$ are strongly and weakly relevant, respectively, with $\mathsf{Corr}([q_{1,t-1}\ q_{2,t-1}]', x^*_{t-1}) = [0.85,\ 0.1]'$. Simulations are based on 200,000 repetitions. All errors are drawn from t-distributions.

		d $\mathsf{Corr}([q_{1,t-1}\ q_{2,t-1}]', \xi_t)$ β																										
		0.1									0.295									0.49								
		$[0.5,\ -0.6]'$			$[0,\ 0]'$			$[-0.4,\ 0.3]'$			$[0.5,\ -0.6]'$			$[0,\ 0]'$			$[-0.4,\ 0.3]'$			$[0.5,\ -0.6]'$			$[0,\ 0]'$			$[-0.4,\ 0.3]'$		
$\sigma_\xi, \sigma_\varepsilon, \sigma_\eta, \sigma_\mu$	T	-2	0	3	-2	0	3	-2	0	3	-2	0	3	-2	0	3	-2	0	3	-2	0	3	-2	0	3	-2	0	3
1.7,1.7,1.7,1.7	250	99	100	76	5.3	5.1	5.3	52	97	55	98	100	71	5.3	5.1	5.3	48	97	51	94	100	57	5.4	5.1	5.4	37	95	40
1.7,1.7,1.7,1.7	1000	100	100	99	5.1	5.0	5.1	96	100	96	100	100	99	5.1	5.0	5.0	93	100	93	100	100	94	5.1	4.9	5.0	78	100	80
1.7,1.7,1.7,1.1	250	99	100	75	5.2	5.0	5.3	49	98	52	98	100	70	5.2	5.0	5.3	45	97	48	94	100	55	5.3	5.0	5.4	35	95	38
1.7,1.7,1.7,1.1	1000	100	100	100	5.1	5.0	5.0	96	100	96	100	100	99	5.1	4.9	5.1	93	100	93	100	100	94	5.2	4.9	5.2	77	100	79
1.7,1.7,1.1,1.7	250	100	100	94	5.1	5.1	5.2	74	98	75	100	100	92	5.1	5.2	5.2	70	98	71	99	100	82	5.2	5.1	5.2	57	97	60
1.7,1.7,1.1,1.7	1000	100	100	100	5.1	4.9	5.1	100	100	100	100	100	100	5.0	4.9	5.1	99	100	99	100	100	100	5.1	5.0	5.1	96	100	96
1.7,1.7,1.1,1.1	250	100	100	94	5.2	5.1	5.2	71	99	73	100	100	91	5.2	5.0	5.2	67	99	69	100	100	80	5.1	5.1	5.3	54	98	57
1.7,1.7,1.1,1.1	1000	100	100	100	5.1	5.0	5.1	100	100	100	100	100	100	5.0	5.0	5.1	100	100	100	100	100	100	5.0	5.1	5.1	96	100	96
1.7,1.1,1.7,1.7	250	99	100	73	5.5	5.0	5.5	48	95	59	99	100	68	5.5	5.0	5.5	44	94	55	95	99	54	5.7	4.9	5.9	35	90	45
1.7,1.1,1.7,1.7	1000	100	100	99	5.1	4.9	5.1	93	100	97	100	100	99	5.1	5.0	5.1	90	100	96	100	100	92	5.2	4.9	5.2	72	100	85
1.7,1.1,1.7,1.1	250	99	100	71	5.4	5.0	5.4	45	95	57	99	100	66	5.4	5.0	5.4	41	94	53	95	100	52	5.7	4.9	5.8	32	90	43
1.7,1.1,1.7,1.1	1000	100	100	99	5.1	5.0	5.1	93	100	97	100	100	99	5.1	5.0	5.1	89	100	96	100	100	92	5.2	5.0	5.2	71	100	84
1.7,1.1,1.1,1.7	250	100	100	93	5.1	5.0	5.3	70	98	79	100	100	90	5.1	5.1	5.3	66	98	75	100	100	78	5.3	5.0	5.3	53	96	64
1.7,1.1,1.1,1.7	1000	100	100	100	5.1	5.0	5.1	99	100	100	100	100	100	5.0	5.1	5.1	99	100	100	100	100	99	5.1	5.0	5.0	93	100	98
1.7,1.1,1.1,1.1	250	100	100	92	5.1	5.0	5.2	67	98	77	100	100	89	5.1	5.0	5.2	63	98	73	100	100	77	5.3	5.0	5.4	50	97	62
1.7,1.1,1.1,1.1	1000	100	100	100	5.1	5.0	5.0	100	100	100	100	100	100	5.1	5.1	5.1	99	100	100	100	100	100	5.1	5.0	5.0	94	100	98
1.1,1.7,1.7,1.7	250	94	100	56	5.2	5.1	5.2	37	98	32	92	100	50	5.3	5.1	5.3	33	97	29	81	100	37	5.5	5.0	5.5	25	95	22
1.1,1.7,1.7,1.7	1000	100	100	97	5.1	5.1	5.0	85	100	79	100	100	94	5.0	5.1	5.0	79	100	72	99	100	79	5.1	5.0	5.1	57	100	50
1.1,1.7,1.7,1.1	250	94	100	54	5.2	5.1	5.2	34	98	30	92	100	49	5.2	5.0	5.3	31	98	27	81	100	36	5.4	4.9	5.4	23	95	21
1.1,1.7,1.7,1.1	1000	100	100	97	5.1	5.1	5.1	84	100	78	100	100	94	5.2	5.0	5.1	78	100	70	99	100	78	5.1	5.0	5.2	56	100	49
1.1,1.7,1.1,1.7	250	100	100	83	5.2	5.1	5.2	59	99	54	99	100	78	5.1	5.1	5.2	54	99	49	97	100	62	5.1	5.0	5.2	42	98	37
1.1,1.7,1.1,1.7	1000	100	100	100	5.0	5.0	5.0	98	100	97	100	100	100	5.1	5.0	5.0	97	100	95	100	100	97	5.0	5.0	4.9	86	100	80
1.1,1.7,1.1,1.1	250	100	100	82	5.1	5.2	5.2	56	99	51	100	100	77	5.1	5.1	5.2	51	99	46	97	100	60	5.3	5.1	5.3	39	99	34
1.1,1.7,1.1,1.1	1000	100	100	100	5.0	5.0	5.0	99	100	97	100	100	100	5.0	5.1	5.0	97	100	95	100	100	98	5.1	5.0	5.1	85	100	79
1.1,1.1,1.7,1.7	250	95	100	54	5.5	4.9	5.4	34	96	35	93	100	48	5.4	4.9	5.5	31	95	32	83	99	36	5.7	4.9	5.8	24	91	24
1.1,1.1,1.7,1.7	1000	100	100	96	5.1	5.1	5.0	81	100	83	100	100	93	5.1	5.0	5.1	74	100	77	100	100	75	5.3	4.9	5.2	52	100	55
1.1,1.1,1.7,1.1	250	96	100	51	5.4	4.9	5.5	32	96	33	93	100	46	5.5	4.9	5.6	29	95	30	83	100	35	5.7	5.0	5.8	22	90	23
1.1,1.1,1.7,1.1	1000	100	100	96	5.0	5.0	5.0	80	100	82	100	100	92	5.1	4.9	5.1	73	100	75	100	100	74	5.2	4.9	5.2	51	100	53
1.1,1.1,1.1,1.7	250	100	100	81	5.3	5.1	5.3	56	98	57	100	100	76	5.3	5.1	5.3	51	98	52	98	100	60	5.4	5.1	5.4	39	97	40
1.1,1.1,1.1,1.7	1000	100	100	100	5.1	5.0	5.1	97	100	98	100	100	100	5.1	5.0	5.1	95	100	96	100	100	96	5.0	5.0	5.1	81	100	84
1.1,1.1,1.1,1.1	250	100	100	80	5.2	5.0	5.2	52	99	54	100	100	74	5.2	5.0	5.2	48	99	49	98	100	58	5.3	5.1	5.3	36	98	37
1.1,1.1,1.1,1.1	1000	100	100	100	5.0	5.0	5.0	98	100	98	100	100	100	5.0	4.9	5.1	96	100	97	100	100	96	5.0	4.9	5.1	81	100	83

Table 6: Long-Memory Estimates

The upper panel of the table reports estimates of d using the multivariate EW estimator of Nielsen and Shimotsu (2007) for $Y_t = [rv_t,\ bv_t,\ vix_t^2,\ r_t]'$. The size of the spectral window is set to $m = T^{0.35}$; the choice is based on a graphical analysis of the slope of the log periodograms as suggested by Beran (1994). $t_{d=0}$ denotes the respective t-statistic of element i of Y_t given by $2\sqrt{m}\hat{d}_i$. The lower panel of the table summarizes the t-statistics corresponding to the null hypothesis $d_i = d_j$ for $i \neq j$. Nielsen and Shimotsu (2007) define the t-statistic as

$$t_{d_i=d_j} = \frac{\sqrt{m}\left(\hat{d}_i - \hat{d}_j\right)}{\sqrt{\frac{1}{2}\left(1 - \frac{\hat{\tau}_{i,j}^2}{\hat{\tau}_{i,i}\hat{\tau}_{j,j}}\right) + h(T)}},$$

where $\hat{\tau}_{i,j} = \frac{1}{m}\sum_{l=1}^{m} \text{real}\{I(\lambda_l)\}$ and $I(\lambda_l)$ is the periodogram of a (4×1) vector with elements $\Delta^{\hat{d}_i} Y_{t,i}$ at frequency λ_l. $h(T)$ is a tuning parameter, which we set equal to $(\ln(T))^{-1}$ as in Nielsen and Shimotsu (2007). The resulting statistic $t_{d_i=d_j}$ should be compared to critical values from a t-distribution.

	Estimates for d			
	rv_t	bv_t	vix_t^2	r_t
$\hat{d}$	0.3517	0.3403	0.4393	0.0202
$t_{d=0}$	2.8134	2.7227	3.5146	0.1618
	$t_{d_i=d_j}$ statistics with $h(T) = 0.1233$			
	rv_t	bv_t	vix_t^2	r_t
rv_t	-	0.2609	-1.6730	2.2602
bv_t		-	-1.7115	2.1587
vix_t^2			-	2.9844
r_t				-

Table 7: Summary Statistics and Estimation Results

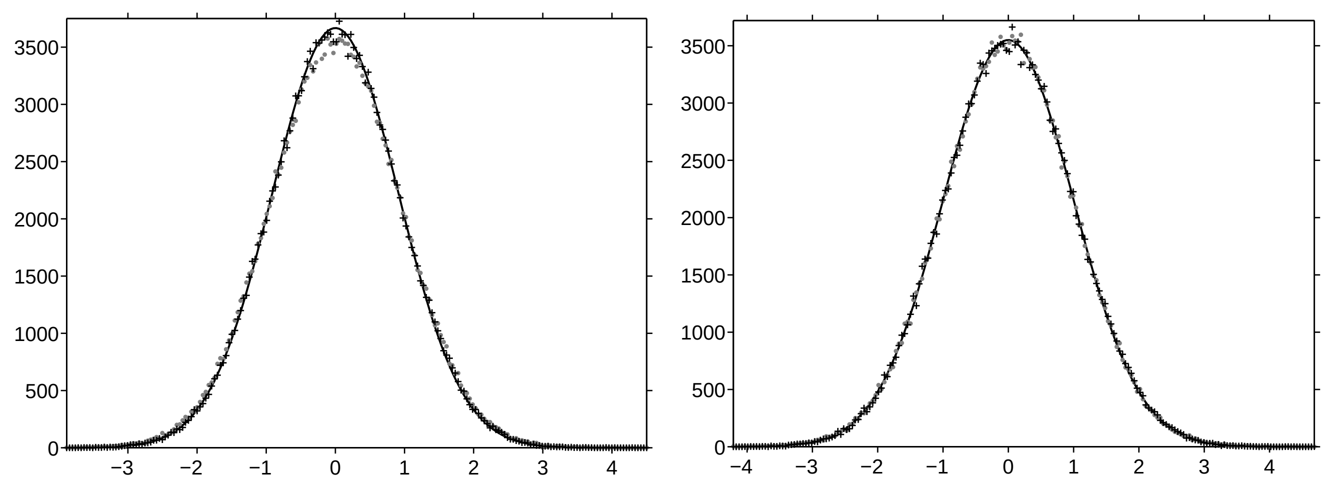
The first panel of the table reports summary statistics of the three variance series and intraday returns. The second panel summarizes the estimation results when the predictive regression (29) is evaluated by OLS. OLS-SE denotes the usual standard error of b, and HAC-SE reports standard errors based on HAC covariance estimation using a Bartlett kernel. The third panel of the table contains the analogous results from IV estimation. IV-SE is the usual standard error of b and $\mathcal{J}$ is Sargan's statistic from Corollary 1.

	Summary Statistics					
			Autocorrelation			
	Average	Std. Dev.	1	2	3	22
r_t	0.0139	1.2833	-0.0769	-0.0612	0.0205	0.0356
rv_t	31.7888	47.6128	0.9976	0.9927	0.9860	0.7009
bv_t	25.4352	40.5219	0.9976	0.9927	0.9858	0.6959
vix_t^2	45.9689	48.6179	0.9690	0.9469	0.9322	0.7413
	OLS Regressions (29)					
x_t	$\hat{b}$	OLS-SE(b)	HAC-SE(b)			

Table 7 – continued from previous page

rv_t	$4.98{\times}10^{-5}$	0.0005	0.0004	
bv_t	$7.70{\times}10^{-5}$	0.0005	0.0004	
vix_t^2	0.0015	0.0005	0.0002	
		`IV Regressions` (29)		
x_t	$\hat{b}$	IV-SE(b)	HAC-SE(b)	$\mathcal{J}$
rv_t	-0.0130	0.0024	0.0046	13.7306
bv_t	-0.0088	0.0021	0.0035	30.3321
vix_t^2	0.0128	0.0020	0.0060	1.4128

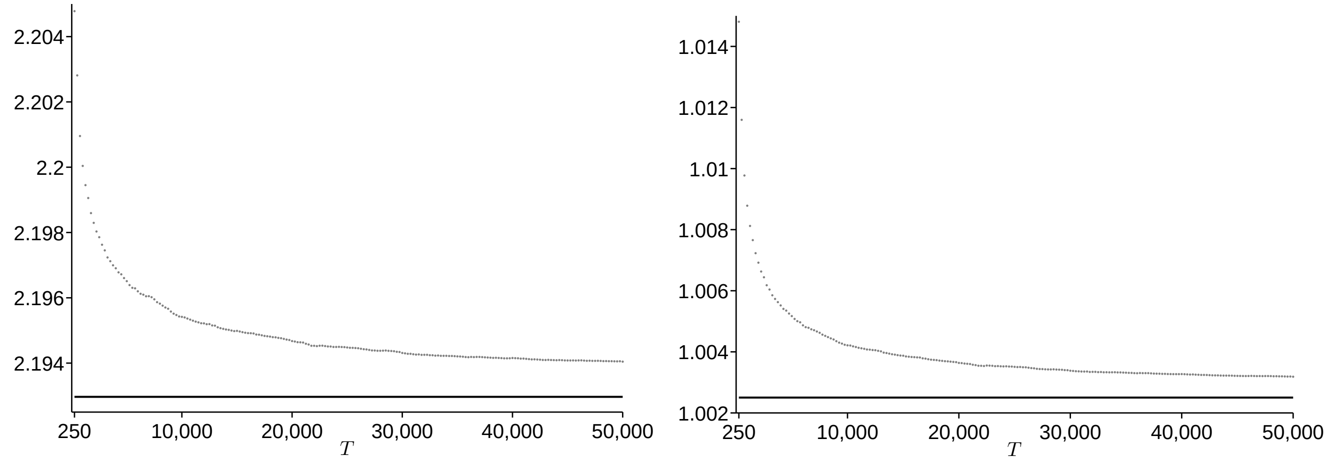
Figures



(i) OLS: Distribution of $T^{1/2}\hat{b}$

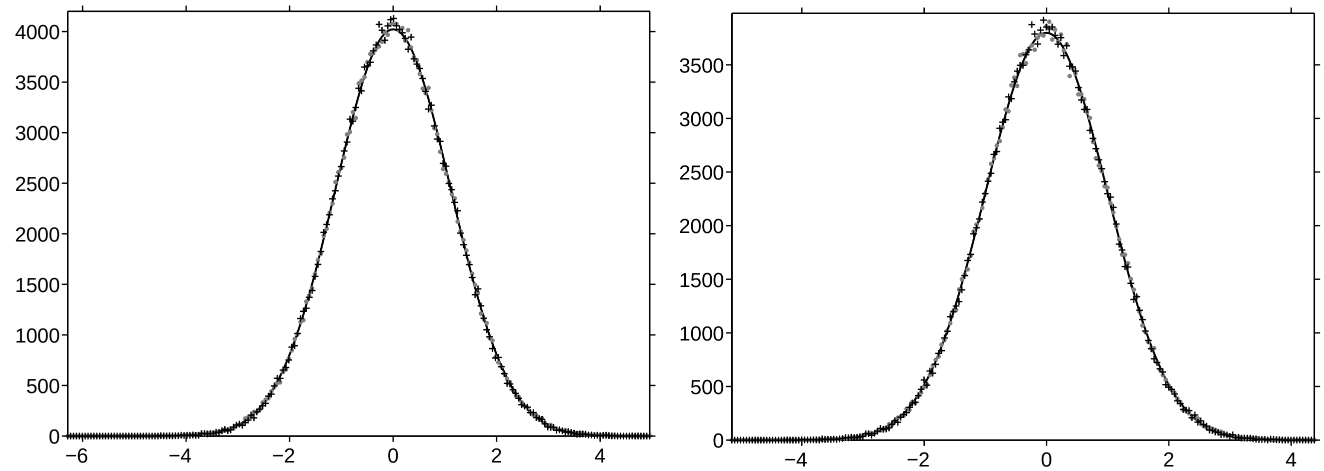
(ii) OLS: Distribution of t_b

Figure 1: **Small sample behavior of OLS estimates if** $\beta = 0$ - The figures plot the small sample distribution of the scaled OLS estimate $T^{1/2}\hat{b}$ from 200,000 simulations of the DGP (2)-(5), and the associated t-statistic, t_b. The black solid line reports the asymptotic distribution from Theorem 1. The gray dots represent the empirical distribution for $T = 250$; the black crosses are the empirical distribution for $T = 50,000$. In the simulations, we let $d = 0.35$, $\sigma_\eta \approx 1$, $\sigma_\xi \approx 2$, $\sigma_\varepsilon \approx 1.8$, $\alpha = 1.2$, and $\beta = 0$. The innovations in the DGP are drawn from continuous uniform distributions.



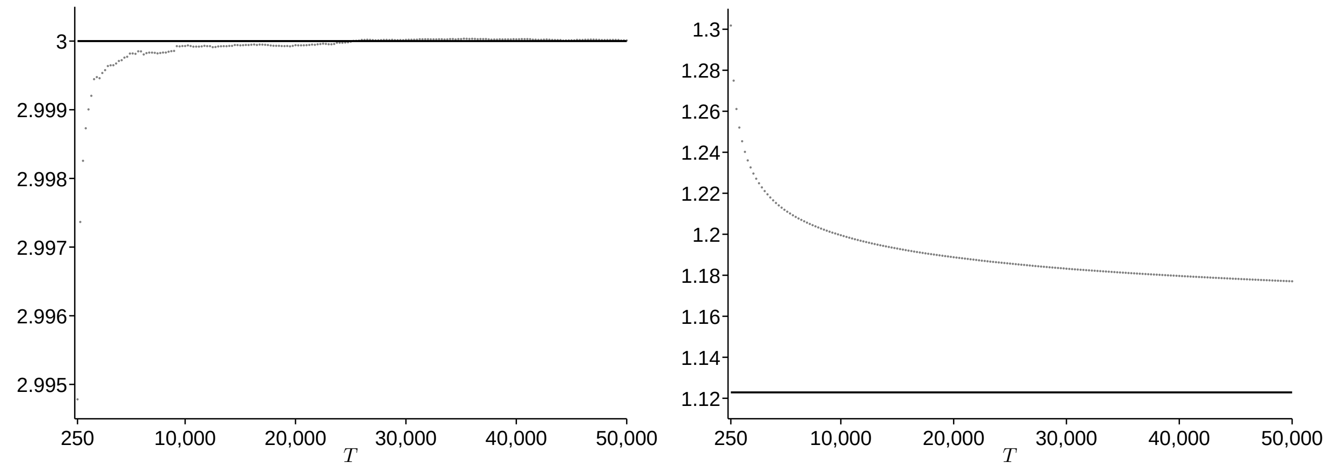
(i) OLS: $\hat{b}$ (ii) OLS: $T^{-1/2}t_b$

Figure 2: **Small sample behavior of OLS estimates if** $\beta \neq 0$ - The figures plot the small sample behavior of the OLS estimate $\hat{b}$ from 200,000 simulations of the DGP (2)-(5), and the associated scaled t-statistic, $T^{-1/2}t_b$. The x-axis contains varying sample sizes from $T = 250$ to $T = 50,000$. The black solid line reports the asymptotic value from Theorem 1. The gray dots represent the average estimate for a given T. In the simulations, we let $d = 0.2$, $\sigma_\eta \approx 1.2$, $\sigma_\xi \approx 1.7$, $\sigma_\varepsilon \approx 1.4$, $\alpha = 1.2$, and $\beta = 4$. The innovations in the DGP are drawn from t-distributions.



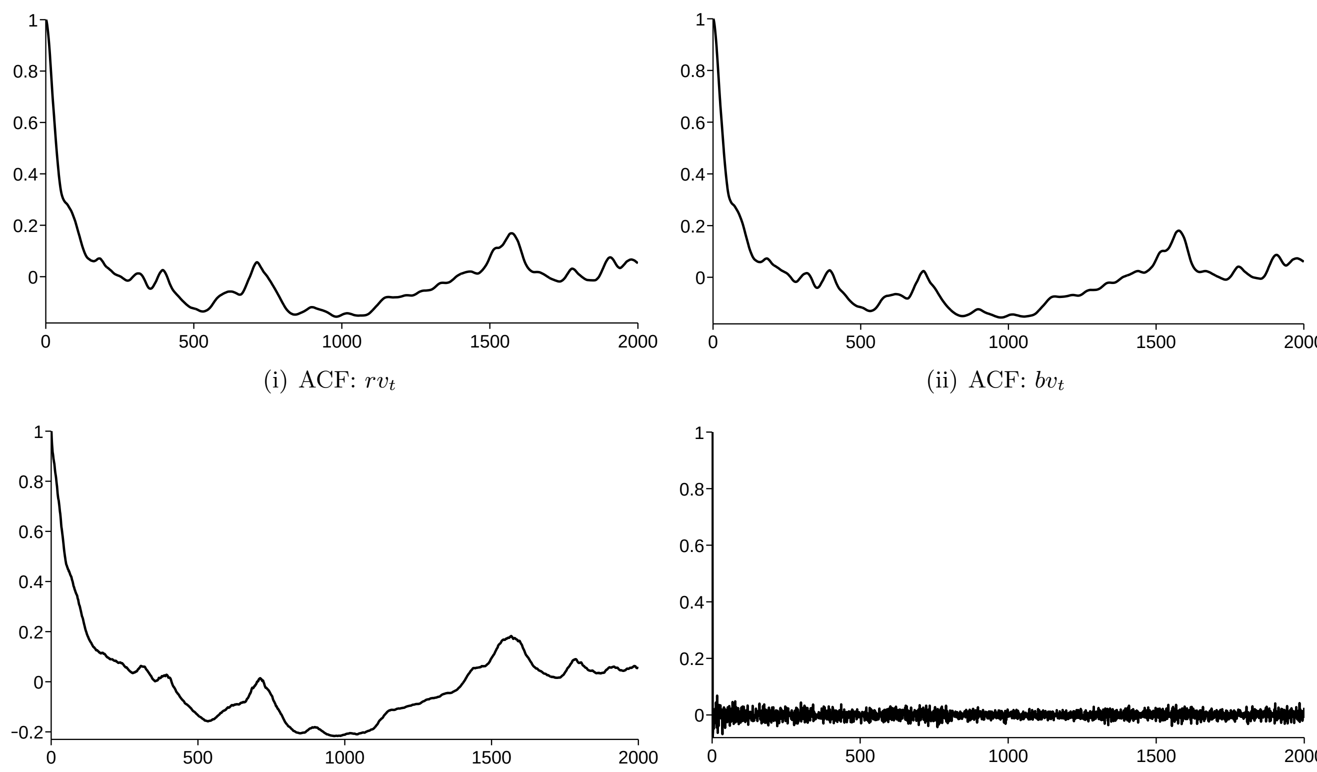
(i) IV: Distribution of $T^{1/2}\hat{b}$ (ii) IV: Distribution of t_b

Figure 3: **Small sample behavior of IV estimates if** $\beta = 0$ - The figures plot the small sample distribution of the scaled IV estimate $T^{1/2}\hat{b}$ from 200,000 simulations of the DGP (2)-(5) and the instruments (19) with $K = 9$, and the associated t-statistic, t_b. The black solid line reports the asymptotic distribution from Theorem 2. The gray dots represent the empirical distribution for $T = 250$; the black crosses are the empirical distribution for $T = 50,000$. In the simulations, we let $d = 0.3$, $\sigma_\eta = 1$, $\sigma_\xi = 2$, $\sigma_\varepsilon = 1.8$, $\sigma_\upsilon = [1.5, 1.2, 3.0, 1.5, 1.2, 3.0, 1.5, 1.2, 3.0]'$, $\alpha = 1.2$, $\beta = 0$, and $\rho = [3.77, 4.44, 1.77, 0.55, 3.11, 2.66, 1.99, 3.99, 0.99]'$. The innovations in the DGP are drawn from standard normal distributions.



(i) IV: $\hat{b}$ (ii) IV: $T^{-1/2}t_b$

Figure 4: **Small sample behavior of IV estimates if** $\beta \neq 0$ - The figures plot the small sample behavior of the IV estimate $\hat{b}$ from 200,000 simulations of the DGP (2)-(5) and the instruments (19) with $K = 3$, and the associated scaled t-statistic, $T^{-1/2}t_b$. The x-axis contains varying sample sizes from $T = 250$ to $T = 50,000$. The black solid line reports the asymptotic value from Theorem 2. The gray dots represent the average estimate for a given T. In the simulations, we let $d = 0.4$, $\sigma_\eta \approx 1.0$, $\sigma_\xi \approx 2.0$, $\sigma_\varepsilon \approx 1.8$, $\sigma_v \approx [1.5, 1.2, 3.0]'$, $\alpha = 1.2$, $\beta = 3$, and $\rho = [3.77, 4.44, 1.77]'$. The innovations in the DGP are drawn from continuous uniform distributions.



(i) ACF: rv_t (ii) ACF: bv_t

(iii) ACF: vix_t^2 (iv) ACF: r_t

Figure 5: **ACF estimates for the three variance series and returns** - The figure plots the estimates of the autocorrelation of the realized variance, rv_t, the bipower variation, bv_t, the volatility index, vix_t^2, and daily intraday returns on the the S&P 500, r_t. The x-axis measures lags in daily units.

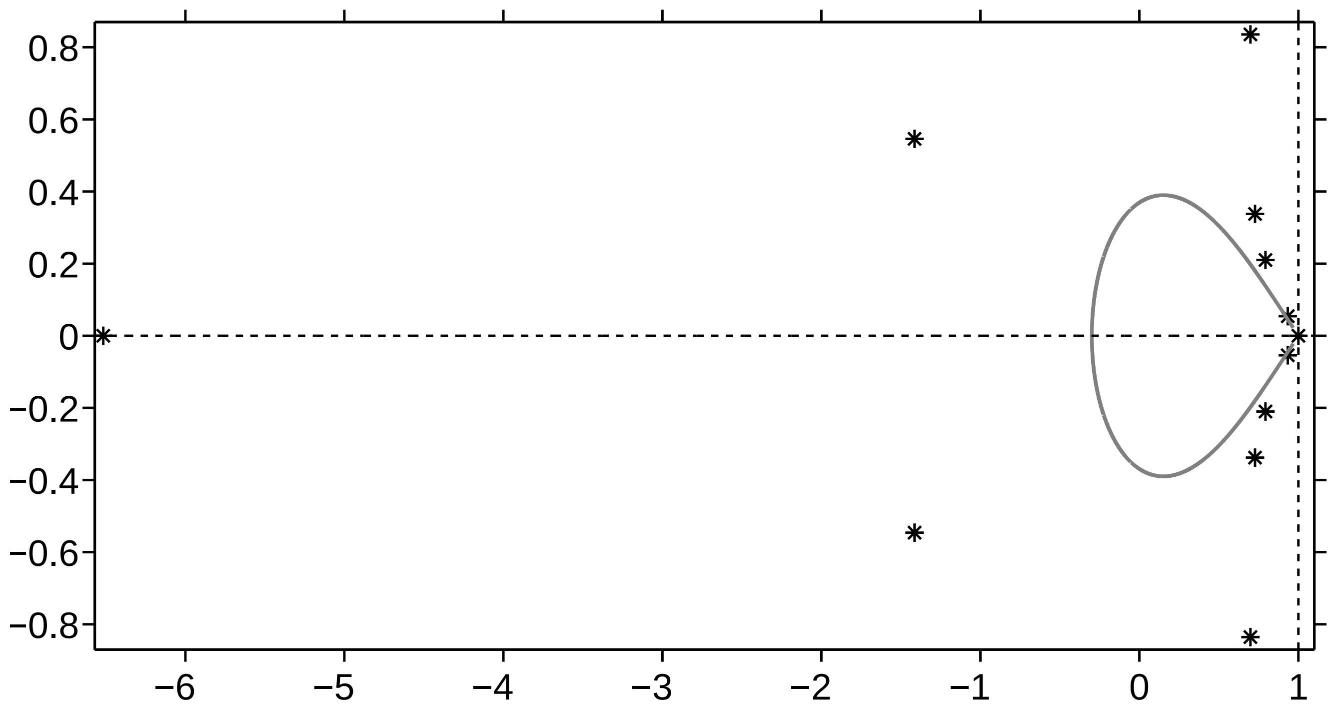


Figure 6: **Roots of the characteristic polynomial of the co-fractional VAR** - The figure plots the roots of the characteristic equation $|(1-c)I_{3\times3} - \varphi\theta' c - (1-c)\sum_{i=1}^{n}\Gamma_i c^i| = 0$, indicated by the black stars. The gray line is the image of the complex disk $\mathbb{C}_d$, for $\hat{d} = 0.3775$. For $\theta' X_t$ to be $I(0)$, all roots must be equal to one or lie outside the disk.