Indirect inference with time series observed with error*

Eduardo Rossi†
University of Pavia
Italy

Paolo Santucci de Magistris
CREATES, Aarhus University
Denmark

December 31, 2014

Abstract

We analyze the properties of the indirect inference estimator when the observed series are contaminated by measurement error. We show that the indirect inference estimates are asymptotically biased when the nuisance parameters of the measurement error distribution are neglected in the indirect estimation. We propose to solve this inconsistency by jointly estimating the nuisance and the structural parameters. Under standard assumptions, this estimator is consistent and asymptotically normal. A condition for the identification of ARMA plus noise is obtained. The proposed methodology is used to estimate the parameters of continuous-time stochastic volatility models with auxiliary specifications based on realized volatility measures. Monte Carlo simulations shows the bias reduction of the indirect estimates obtained when the microstructure noise is explicitly modeled. Finally, an empirical application illustrates the relevance of a realistic specification of the microstructure noise distribution to match the features of the observed log-returns at high frequencies.

Keywords: Indirect inference, measurement error, stochastic volatility, realized volatility

J.E.L. classification: C13, C15, C22, C58

*We are grateful to Torben Andersen, Roberto Renò, Viktor Todorov and the participants to the 2014 Conference on Indirect Estimation Methods in Finance and Economics (Constance) for useful comments on previous versions of this paper. Eduardo Rossi acknowledges financial support from the MIUR PRIN project MISURA - Multivariate Statistical Models for Risk Assessment. Paolo Santucci de Magistris acknowledges financial support from CREATES - Center for Research in Econometric Analysis of Time Series, funded by the Danish National Research Foundation (DNRF78).

†Dipartimento di scienze economiche e aziendali. Via San Felice 5, 27100 Pavia, Italy. Tel.: +39 0382/986207 Fax: +39 0382/304226. E-mail: erossi@eco.unipv.it.
1 Introduction

A common feature of many economic and financial time series is that they are recorded with errors or frictions. In some cases they are estimated rather than observed exactly. For instance, in macroeconomics the error in the measurement of GDP is a well-known problem, see Aruoba et al. (2013). In finance, asset returns sampled at high frequency are subject to a vast number of frictions, so that the observed transaction log-price is the sum of an unobservable efficient price and a noise component due to the imperfections of the trading process. If the observed time series is contaminated by noise then the true relationship between variables is somewhat obscured, with consequences on the estimates of the structural parameters. Well known econometric methods, such as Kalman filter or the instrumental variables, are generally employed to deal with the measurement error in a time series framework and to provide parameter estimates that are consistent also in presence of errors-in-variables. However there are limits to the range of applicability of the standard methodologies. For example, the Kalman filter can only be adopted in a linear-Gaussian state-space framework, while the instrumental variables, when represented by the lagged observed series, are very weak when the signal is not persistent (see Hansen and Lunde, 2014).

We propose an alternative and general methodology, based on indirect inference, to deal with the errors-in-variables problem also when the likelihood function (or any other criterion function that might form the basis of estimation) is analytically intractable or too difficult to evaluate. The indirect inference method has been introduced in the econometric literature by Smith (1993), Gouriéroux et al. (1993), Bansal et al. (1995) and Gallant and Tauchen (1996), and is surveyed in Gouriéroux and Monfort (1996) and Jiang and Turnbull (2004). The estimation consists of two stages. First, an auxiliary statistic is calculated from the observed data. Then an analytical or simulated mapping, called binding function, of the structural parameters to the auxiliary statistic is calculated. Indirect inference chooses the parameters of the economic model so that these two estimates of the parameters of the auxiliary model are as close as possible. The indirect inference estimators are typically placed into one of two categories: score-based estimators made popular by Gallant and Tauchen (1996), or distance-based estimators proposed by Smith (1993) and refined by Gouriéroux et al. (1993). The simulated score-based estimators have the computational advantage that the auxiliary parameters are estimated from the observed data only once. On the other hand, the distance-based estimators must re-estimate the auxiliary parameters from simulated data at each step of the optimization algorithm. The indirect inference methods have been successfully employed in the estimation of continuous-time models for asset prices and volatility, that have experienced important developments in the last twenty years, see among others Gallant et al. (1997), Chernov et al. (2003). Since the transition density functions are often unknown and the stochastic volatility (SV) process is not directly observable, the estimation with latent variables can be carried out using simulation methods.

In this article, we first formally study how neglecting the presence of measurement error in the observed time series affects the indirect inference estimates. Theoretical results, supported by several examples based on ARMA and continuous time stochastic processes, establish that ignoring such contamination makes the indirect inference estimates inconsistent and biased in finite samples. Indeed, as noted by Ghysels and Khalaf (2003) and Dridi et al. (2007) indirect inference theory, as originally proposed by Gouriéroux et al. (1993), Gouriéroux and Monfort (1996) and Gallant and Tauchen (1996), does not take nuisance parameters formally into consideration.

Second, we propose a simple solution to the inconsistency of the indirect inference estimator that is to explicitly account for the presence of the measurement error, and to treat it as a structural feature. This means that the nuisance parameters characterizing the conditional distribution of the noise must be estimated with the structural parameters. This implies that the simulated trajectories of the structural model must be contaminated by measurement error. It follows that the simulated binding function will explicitly depend on the noise parameters, thus leading to a consistent matching of the auxiliary estimates. The main advantage of this approach is that it is fully built within the indirect inference framework, so that the asymptotic consistency and
normality of this estimator can be easily proved under correct specification of the contaminating processes.

Third, the issue of identification is discussed in detail. Indeed, while the indirect inference framework provides a general setup to tackle the problem of measurement errors, choosing an auxiliary model able to identify both the structural and the noise parameters, may be non trivial. A crucial assumption in the indirect inference framework is that the binding function must be locally injective to guarantee identification. We prove that when the structural model is an ARMA\((r,l)\) with \(r > l\) contaminated by i.i.d. noise with non-zero variance, the Jacobian of the binding function generated by autoregressive auxiliary models with \(m > r + l\) lags has full-column rank. Since many economic and financial models have an ARMA representation, this result gives a necessary condition for the identifiability of the structural models in presence of measurement error. As a consequence, when the structural model has an MA(1) plus noise representation, then the identification condition is not satisfied by any auxiliary model. On the contrary, we show that when the discretized trajectories of an Ornstein-Uhlenbeck process are contaminated by an additive i.i.d. noise, the Jacobian matrix of the binding function with respect to all structural and nuisance parameters has full rank.

Fourth, the proposed method is employed in the estimation of continuous-time SV models based on ex-post measurement of daily volatility like realized volatility \((RV)\). When the volatility is generated from an Heston (1993) model, then the binding function relative to the HAR-RV model of Corsi (2009) can be written in terms of the SV parameters so that the identification condition can be verified. We also show by means of Monte Carlo simulations that accounting for the microstructure noise produces unbiased SV parameter estimates, when employing intraday returns at very high frequencies (e.g. 30 seconds). On the contrary, neglecting the microstructure noise leads to severe biases especially in the estimates of the long-run mean of volatility. Alternatively, sampling returns at low intradaily frequencies attenuates the impact of the market microstructure noise on the volatility estimates. In terms of efficiency of the estimates, sampling at the highest frequency has the advantage that no data is discarded. On the other hand, additional nuisance parameters need to be estimated.

Finally, an empirical study based on the RV series of JP Morgan corroborates the evidence emerged in the theoretical study and highlights the advantages and limits of the proposed methodology. In particular, neglecting the microstructure noise makes the estimates of the SV parameters highly dependent on the choice of the sampling frequency adopted in the construction of RV. Indeed, it is not possible to reconcile the estimates of the auxiliary parameters obtained with RV based on log-returns sampled at 5 seconds and 5 minutes, unless the microstructure noise is explicitly modeled. Moreover, using the data at very high frequency allows to verify if the assumed structural model and the generation of the microstructure noise are coherent descriptions of the properties of the observed log-returns. Notably, sampling at low frequencies, as in Corsi and Renò (2012), is sufficient to neutralize the impact of noise on the SV estimates at least when price jumps are excluded from the structural model. However, sampling at low frequencies does not provide sufficient information on the price generation mechanism in a realistic scenario and makes the identification of price jumps very difficult. On the contrary, sampling the returns at 5-seconds allows to evaluate the impact of price decimalization on the indirect inference estimates. It turns out that, when the bipower variation \((BPV)\) sampled at 5-seconds is adopted to disentangle price jumps from volatility dynamics, the fit of the model is dramatically affected, also when a microstructure noise in the form of a bid-ask spread is considered. This calls for a more sophisticated and realistic specification of the microstructure noise.

The paper is organized as follows: in Section 2 we illustrate the effect of the presence of measurement error on the indirect inference estimator in the pure time series case. We formally prove the inconsistency of the indirect inference estimates in this case and we illustrate it by means of several examples. In Section 3 we prove the consistency of the proposed estimator that accounts for the noise and we illustrate its reliability in finite samples by means of a number of example both
in discrete and in continuous time. In Section 4 we illustrate, both theoretically and by means of Monte Carlo simulations, the peculiar problems that arise in the indirect estimation of continuous-time SV model by means of RV, and Section 5 reports the evidence based on real data. Finally, Section 6 concludes.

2 The effect of measurement error on the indirect inference estimator

Following the framework and notation of Gouriéroux et al. (1993), we first present the properties of the indirect inference method in presence of measurement error. The parameters of interest are those in the vector $\theta$ which characterizes the data-generating process of the unobserved series $y_t$. Here we consider the case of a discrete-time process $y_t$ which is contaminated by an error term. To simplify the notation and the exposition of the results, we consider only the dependence on past values of $y_t$, i.e. the pure time series case.

Assumption 1 The process \{$y_t$\} is a strictly stationary and ergodic process with transition density $p(y_t | y_{t-1}; \theta)$, where $y_{t-1} = (y_{t-1}, \ldots, y_{1})$, that is difficult or impossible to evaluate analytically. The vector containing the true structural parameters is $\theta_0 \in \Theta \subseteq \mathbb{R}^p$.

Assumption 2 A sample of $T$ observations \{$x_t\}_{t=1}^T$ is observed as

$$x_t = g(y_t, u_t) \quad t = 1, 2, \ldots, T$$

Assumption 3 The term $u_t$ is the measurement error which is supposed to be covariance stationary with a known conditional distribution, i.e. $f(u_t | u_{t-1}, u_{t-2}, \ldots; \psi_0)$, where $\psi_0 \in \Psi \subseteq \mathbb{R}^h$.

Assumption 1 is rather standard in this framework, as indirect inference requires the process $y_t$ to be stationary with constant moments to be asymptotically valid. Assumption 2 is quite general, it allows a non-linear mapping between the observed series $x_t$, the signal $y_t$ and the measurement error. Generally $g(\cdot)$ is a linear/additive function or can be reduced to be linear, i.e. $x_t = y_t + u_t$. For example, suppose that the observed stock price $P_t$ is equal to a latent efficient price times an error term with positive support, $P_t = P_t^* \cdot \epsilon_t$, then the efficient log-price, $P_t^*$, is contaminated by an additive measurement error term, i.e. $P_t = P_t^* + \epsilon_t$, where $\epsilon_t = \log(\tilde{\epsilon}_t)$. Assumption 3 characterizes the dynamic features of the measurement error, which depends on a number of true nuisance parameters, contained in the vector $\psi_0$, and it does not exclude correlation between the signal and the noise and autocorrelation in the measurement error. As it will be clear from the discussion in Section 3, the knowledge of the functional form of the conditional distribution of $u_t$ is crucial when we want to simulate from it.

We now investigate the impact of neglecting the possible presence of measurement error when carrying out indirect inference on $\theta_0$ employing observations of the contaminated process $x_t$. The indirect inference consists of two steps: the estimation of the auxiliary (or instrumental) model and the calibration. The auxiliary model is defined by a conditional probability density function $f(x_t | x_{t-1}; \beta)$ which depends on a $q$-dimensional parameter vector, $\beta \in B \subseteq \mathbb{R}^q$. This density has a convenient analytical expression. The number of parameters in the auxiliary model must be at least as large as the number of parameters in the economic model, i.e., $q \geq p$. The auxiliary model is, in general, incorrectly specified, i.e. need not describe accurately the conditional distribution of $x_t$. The parameters of the auxiliary model can be estimated using the observed data by maximizing the log-likelihood function or any other criterion function, $Q_T(x_T; \beta)$, which satisfies some technical assumptions, see Gouriéroux and Monfort (1996, p.85), i.e.

$$\hat{\beta}_T = \arg \max_{\beta} Q_T(x_1, \ldots, x_T; \beta).$$

(2)
In a likelihood setting, identification requires the true densities of the data being "smoothly embedded" within the scores of the auxiliary model, see Gallant and Tauchen (1996). The criterion is assumed to tend asymptotically (and uniformly almost certainly) to a non-stochastic limit (see Gouriéroux et al., 1993, Assumption 2)

$$\lim_{T \to \infty} Q_T(x_1, \ldots, x_T; \beta) = Q_\infty(\theta_0, \psi_0, \beta).$$  \hspace{1cm} (3)

When the series is measured without noise, this limit depends only on the unknown auxiliary parameter $\beta$ and on the true parameter of interest $\theta_0 \in \Theta \subseteq \mathbb{R}^p$. However when the series at hand is contaminated by noise this limit depends also on the true nuisance parameter vector, $\psi_0 \in \Psi \subseteq \mathbb{R}^h$, where $h$ is the dimension of $\psi_0$. For example, when $u_t$ is assumed to be an i.i.d.$N(0, \sigma^2_u)$, then $\psi_0 = \sigma^2_u, 0$ and $h = 1$. As in Gouriéroux et al. (1993) we assume that this limit criterion is continuous in $\beta$ and has a unique maximum $\beta_0 = \arg \max_{\beta \in B} Q_\infty(\theta_0, \psi_0, \beta)$.

The binding function, i.e. the link between the auxiliary model parameters and the structural parameters, is given by

$$b(\theta, \psi) = \arg \max_{\beta \in B} Q_\infty(\theta, \psi, \beta)$$  \hspace{1cm} (4)

it follows that $\beta_0 = b(\theta_0, \psi_0)$.

$\hat{\beta}_T$ is a consistent estimator of $b(\theta_0, \psi_0)$ which is an unknown function that depends on $\theta_0$ and $\psi_0$.

In the second step of the procedure, we simulate $S$ trajectories from the DGP of $y_t$ and the estimation of the auxiliary model is carried out on each simulated series. The corresponding estimator based on the $s$-th simulated path of the signal’s DGP for some $\theta$ is

$$\hat{\beta}_T^s(\theta) = \arg \min_{\beta \in B} Q_T(y_{1}^{(s)}(\theta), \ldots, y_{T}^{(s)}(\theta); \beta).$$

When $T \to \infty$, $\hat{\beta}_T^s(\theta)$ converges to the solution of the limit problem

$$\tilde{b}(\theta) = \arg \max_{\beta \in B} Q_\infty(\theta, \beta)$$

that is

$$\lim_{T \to \infty} \hat{\beta}_T^s(\theta) = \tilde{b}(\theta)$$

therefore $\hat{\beta}_T^s(\cdot)$ is an inconsistent functional estimator of $b(\theta, \psi)$. The consequence is that the indirect inference estimator of $\theta$, found by minimizing the distance between $\beta_T$ and $\frac{1}{S} \sum_{s=1}^S \beta_T^s(\theta)$ under a metric given by the positive definite matrix $\Omega$, as

$$\hat{\theta}_{ST}(\psi) = \arg \min_{\theta} \left\| \beta_T - \frac{1}{S} \sum_{s=1}^S \beta_T^s(\theta) \right\|^2_\Omega$$

is inconsistent. The estimator $\hat{\theta}_{ST}(\psi)$ depends on the data via $\hat{\beta}_T$, and thus on the nuisance parameter $\psi$. Indeed, the limit of $\hat{\beta}_T$ as $T \to \infty$ is $b(\theta_0, \psi_0)$. Instead, when $S \to \infty$, $\frac{1}{S} \sum_{s=1}^S \beta_T^s(\theta) \xrightarrow{p} E[\beta_T^s(\theta)]$ and $E[\beta_T^s(\theta)] = \tilde{b}(\theta)$. This discrepancy induces an asymptotic bias in the indirect inference estimator. This can be explicitly seen by using the second order expansion of $\hat{\beta}_T$ in (Gouriéroux and Monfort, 1996, p.77), under the assumption that $q = p$. 

5
Proposition 2.1 If the auxiliary estimator has the following Edgeworth expansion, given Assumption 2, it follows that

$$\hat{\beta}_T(\theta_0, \psi_0) = b(\theta_0, \psi_0) + \frac{A(v_x; \theta_0, \psi_0)}{\sqrt{T}} + \frac{B(v_x; \theta_0, \psi_0)}{T} + o\left(\frac{1}{T}\right) \quad (5)$$

then the indirect inference estimator can be expressed as

$$\hat{\theta}_{ST}(\psi_0) = \theta_0 + \frac{a^*}{\sqrt{T}} + \frac{b^*}{T} + o\left(\frac{1}{T}\right) \quad (6)$$

where

$$a^* = \sqrt{T} \left[ \frac{\partial b(\theta_0)}{\partial \theta'} \right]^{-1} [b(\theta_0, \psi_0) - \tilde{b}(\theta_0)] + \left[ \frac{\partial b(\theta_0)}{\partial \theta'} \right]^{-1} [A(v_x; \theta_0, \psi_0) - \frac{1}{S} \sum_{s=1}^{S} A(v_s; \theta_0)] \quad (7)$$

and

$$b^* = \left[ \frac{\partial b(\theta_0)}{\partial \theta'} \right]^{-1} \left[ B(v_x; \theta_0, \psi_0) - \frac{1}{S} \sum_{s=1}^{S} B(v_s; \theta_0) \right] - \left[ \frac{\partial b(\theta_0)}{\partial \theta'} \right]^{-1} \left\{ \frac{1}{S} \sum_{s=1}^{S} \frac{\partial A(v_s; \theta_0)}{\partial \theta'} a^* + \frac{1}{2} \left( I_q \otimes a^* \right)' F(\theta) a^* \right\}, \quad (8)$$

where

$$F(\theta) = [F_1(\theta)', \ldots, F_q(\theta)']'$$

with $F_i(\theta) = \frac{\partial b_i(\theta)}{\partial \theta}$, and $b_i(\theta)$ the $i$-th element of $b(\theta)$.

Proof: See Appendix B.1.

The proposition establishes that the indirect inference estimator is biased and inconsistent when the data are contaminated by measurement error. It is clear from (7) that $\lim_{T \to \infty} \hat{\theta}_{ST} = \theta_0 + \frac{1}{\sqrt{T}} [b(\theta_0, \psi_0) - \tilde{b}(\theta_0)]$. The term responsible for the distortion and the inconsistency of $\hat{\theta}_{ST}$, i.e. $\left[ \frac{\partial b(\theta_0)}{\partial \theta'} \right]^{-1} [b(\theta_0, \psi_0) - \tilde{b}(\theta_0)]$, does not vanish asymptotically when $\psi_0$ is present. Despite the focus in this paper is on the measurement error, we stress that the content of Proposition 2.1 is rather broad as it formally shows the impact of any neglected nuisance parameter on the consistency of the indirect estimates of the structural parameters. In a DSGE framework, Dridi et al. (2007) show that the binding function depends also on some pseudo-parameters, i.e. quantities which are known to be poorly related to economic reality, and propose to sequentially calibrating them, a method called partial indirect inference. We instead follow a different strategy and, in Section 3, we show that the measurement error could be considered as a structural feature and hence jointly estimated with the other structural parameters.

In the following subsections we present a few examples where the distortion in the binding function induced by the presence of the measurement error can be computed analytically. In the analysis, the framework consists of a linear measurement equation:

$$x_t = y_t + u_t, \quad t = 1, 2, \ldots, T \quad (9)$$

where the measurement error $u_t$ is supposed to be independent and identically distributed with $\text{Var}[u_t] < \infty$ and $\text{Var}[u_t] = \sigma^2_u$, independent of all leads and lags of $y_t$. 


2.1 ARMA processes

The identifiability and estimation of ARMA processes contaminated by additive i.i.d. noise has been studied in literature, see among others Chanda (1995), Jones (1980), Lee and Shin (1997) and Maraval (1979). More recently, de Luna and Genton (2001) have proposed an indirect estimation method to robustify the estimation of ARMA when outliers contaminates observations of the stochastic process. In this section, we show the consequences for the indirect estimation of ARMA models when the signal is contaminated by measurement error. We derive closed-form expressions of the binding function when an autoregressive auxiliary model is adopted and the ARMA signal is contaminated by an i.i.d. measurement error.

As a first example, we consider a zero-mean stationary ARMA(1, 1) signal \( y_t \)

\[
(1 - \alpha L)y_t = (1 + \varphi \varepsilon_t) \quad \varepsilon_t \sim i.i.d.\ N(0, \sigma^2_{\varepsilon})
\]

The polynomials \( 1 - \alpha z \) and \( 1 + \varphi z \) have roots outside the unit circle. Given the result of Granger and Morris (1976), \( x_t \) is also an ARMA(1, 1). Since the measurement error is uncorrelated with constant variance, the error leads to an increase in the variance of \( \varepsilon_t \). The parameter vector of the ARMA(1, 1) is \( \theta = (\alpha, \varphi, \sigma^2_{\varepsilon})' \). Suppose the auxiliary model is an AR(2) process

\[
(1 - \phi_1 L - \phi_2 L^2)x_t = \varepsilon_t
\]

with parameters in \( \beta = (\phi_1, \phi_2, \sigma^2_{\varepsilon})' \) that can be estimated by OLS. In this case, \( p = q \) and the model is exactly identified. Let \( \phi = (\phi_1, \phi_2)' \), the binding function is

\[
b(\theta, \sigma^2_{\varepsilon}) = p \lim_{T \to \infty} \left[ \frac{\hat{\theta}^T}{\hat{\sigma}^2_{\varepsilon},T} \right]
\]

with

\[
p \lim \hat{\theta} = \frac{(\alpha + \varphi)(1 + \alpha \varphi)}{(1 - \alpha^2)} \left[ \frac{(1 + \alpha^2 + 2(\alpha + \varphi)\sigma^2_{\varepsilon}}{1 - \alpha^2} + \frac{\sigma^2}{\sigma^2_{\varepsilon}} \right] - \frac{\alpha(\alpha + \varphi)(1 + \alpha \varphi)}{1 - \alpha^2}
\]

and

\[
p \lim \hat{\sigma}^2_{\varepsilon,T} = \frac{(\sigma^2_{\varepsilon} + 2\alpha \varphi + \sigma^2)}{1 - \alpha^2} \left[ 1 - \frac{(\sigma^2_{\varepsilon} + 2\alpha \varphi + \sigma^2)}{1 - \alpha^2} \right] + \frac{(\sigma^2 + \alpha(\alpha + \varphi)(1 + \alpha \varphi)}{1 - \alpha^2}
\]

The function \( b(\theta) \), which is the limit of the estimators of the auxiliary parameters based on the simulated data, is simply obtained by setting \( \sigma^2_{\varepsilon} = 0 \) in (11) and (12). It is evident that the inconsistency of \( \hat{\theta}_{ST} \) is caused by the presence of measurement error in the data used to estimate \( \beta \). Hence, the term \( \frac{\alpha}{\sqrt{T}} \) in the expansion (6) does not vanish asymptotically when \( T \to \infty \) as the difference \( b(\theta, \sigma^2_{\varepsilon}) - \hat{b}(\theta) \) is non zero when \( \sigma^2_{\varepsilon} > 0 \).

The estimation of a contaminated MA(1) process is another example of the inconsistency of the indirect inference estimator. The data-generating process of the signal is

\[
y_t = \varepsilon_t + \varphi \varepsilon_{t-1} \quad t = 1, \ldots, T
\]

with \( \varepsilon_t \sim i.i.d.\ N(0, \sigma^2_{\varepsilon}) \). The parameters to be estimated, using the observations on \( x_t \), are \( \theta = (\varphi, \sigma^2_{\varepsilon})' \). The auxiliary model is an AR(1),

\[
x_t = \phi x_{t-1} + \varepsilon_t
\]

which corresponds, ignoring the presence of the measurement error, to an exactly identified case. The limit of the OLS estimates of (14), is

\[
b(\theta, \sigma^2_{\varepsilon}) = \lim_{T \to \infty} \left[ \frac{\sum_{t=1}^{T} \frac{y_{t-1} - x_{t-1}}{x_{t-1}}}{\sum_{t=1}^{T} \varepsilon_t^2} \right] = \left[ \frac{(1 + \varphi^2) + \sigma^2_{\varepsilon}/\sigma^2_{\varepsilon}}{(1 + \varphi^2)\sigma^2_{\varepsilon} + \sigma^4} \right]
\]
with $E[x_t x_{t-1}] = \varphi \sigma_z^2$ and $\text{Var}[x_t] = (1 + \varphi^2)\sigma_z^2 + \sigma_u^2$. Whereby the binding function when $\sigma_u^2 = 0$ would be

$$
\tilde{b}(\theta) = \begin{bmatrix}
\frac{\varphi}{1+\varphi^2} \\
\frac{1-\varphi^2}{1-\varphi^2} \\
\end{bmatrix}.
$$

From (16) it is evident that the measurement error impacts on the binding function also in this case. In fact, the first element of $b(\theta, \sigma_u^2)$ is smaller than the corresponding element in $\tilde{b}(\theta)$. This attenuation bias depends on the ratio $\sigma_u^2/\sigma_z^2$. In this case the dependence of $b(\theta, \sigma_u^2)$ on the variance of the measurement error, $\sigma_u^2$, makes the indirect inference estimator of $\varphi$ and $\sigma_u^2$ inconsistent. In general, indirect inference corrects for the inconsistency caused by the misspecification of the auxiliary model. However, in this case, for any $\theta_0 = (\varphi_0, \sigma^2_{z,0})$ and $\psi_0 = \sigma^2_{u,0}$, the term $b(\theta_0, \psi_0) - \tilde{b}(\theta_0)$, which enters in the Edgeworth expansion (7), does not vanish, thus inducing an asymptotic bias in $\theta_{ST}$. Interestingly, $b(\theta_0, \psi_0) - \tilde{b}(\theta_0)$ is null only when $\sigma^2_{u,0} = 0$, i.e. when measurement error is absent.

### 2.2 Continuous-time models

Let suppose that we are interested in estimating the parameter vector $\theta_0$ which characterizes now the transition density, often unknown, of a continuous-time process $\{z(t)\}$. Denote with $y_t$ the function of discrete observations on $z(t)$

$$
y_t = h(z(t); \Delta), \quad t = 1, 2, \ldots, T
$$

where $h(\cdot; \Delta)$ is a known function and $\Delta$ is a known parameter which represents the discretization step. In a simple case, the series $y_t$ can be the discretized $z(t)$ process. It is well known that when $y_t$ is observed in place of $z(t)$ standard indirect inference procedures corrects for the discretization error, see Gouriéroux et al. (1993). However, there are cases where transformations of discrete realizations of the process $z(t)$, employed in the indirect estimation of continuous-time models, are likely to be contaminated by measurement error. In other words, the observed process $x_t$ is the result of the interaction between the latent signal, $z(t)$, the measurement error, $u_t$, the function $h(\cdot; \Delta)$ and the discretization step, $\Delta$. Hence, $x_t$ is

$$
x_t = g[h(z(t); \Delta), u_t], \quad t = 1, 2, \ldots, T
$$

In cases like this, it can be impossible to obtain a closed form expression of the likelihood function for the parameters which characterize the process $z(t)$ and $u_t$ based only on the observation of $x_t$. The examples presented below illustrate the impact that measurement error has on the indirect inference estimator.

### 2.2.1 Geometric Brownian Motion

Consider the geometric Brownian motion (gBm) with drift:

$$
z(t) = \mu z(t)dt + \sigma z(t)dW(t)
$$

(17)

where $\mu$ and $\sigma$ are the drift and volatility parameters and $W(t)$ is a is a standard Brownian motion on $\mathbb{R}$. Let $\theta = (\mu, \sigma)'$ is the vector of structural parameters. The initial value of $z(0)$ is a given random variable (possibly, a constant) taken to be independent of $\{W(t)\}_{t \geq 0}$. The data $z(t)$ is recorded discretely at points $(\Delta, 2\Delta, \ldots, n\Delta)$ in the time interval $[t-1, t]$ with $t = 1, \ldots, T$, that is $z_{t-1 + i\Delta}$ for $i = 1, \ldots, n = 1/\Delta$. Let assume without loss of generality that $n = 1$ so that $y_t = z_t$ is the discrete realizations of $z(t)$ on the unit interval. We assume $y_t$ is contaminated by
$u_t \sim i.i.d.(0, \sigma_u^2)$, independent of all lags and leads of $y_t$. Similarly to Gouriéroux and Monfort (1996), let the auxiliary model be:

$$x_t = x_{t-1} + \phi x_{t-1} + x_{t-1} e_t,$$

with $e_t \sim i.i.d.N(0, \sigma^2)$. The maximum likelihood (ML) estimators of $\phi$ and $\sigma^2$ are

$$\hat{\phi}_T = \left[ \frac{\hat{\phi}_T}{\hat{\sigma}_{e,T}^2} \right] = \left[ \frac{1}{T} \sum_{t=1}^{T} (w_t - \bar{w}_T)^2 \right]$$

(18)

where $w_t = \frac{x_t}{x_{t-1}} = \frac{y_{t-1} + u_{t-1}}{y_{t-1} + u_{t-1}} = \frac{y_{t-1}}{y_{t-1} + u_{t-1}} + \frac{u_{t-1}}{y_{t-1} + u_{t-1}}$. To obtain an expression for the first element of the binding function, i.e. $b_1(\theta, \sigma_u^2)$, we can employ the Edgeworth expansion in (5) (see Genton and Ronchetti, 2003):

$$\hat{\phi}_T = [E[w_t] - 1] + [\bar{w}_T - E[w_t]] = b_1(\theta, \sigma_u^2) + \frac{1}{\sqrt{T}} A(\mu).$$

(19)

Given that $E[\frac{w_t}{y_{t-1} + u_{t-1}}] = 0$ by the independence assumption, and exploiting the fact that the conditional expectation of $\frac{w_t}{y_{t-1}}$ is independent of $x_{t-1}$, the binding function for $\hat{\phi}_T$ is given by

$$b_1(\theta, \sigma_u^2) = E[w_t] - 1 = e^\mu \left[ 1 + E\left( \frac{u_{t-1}^2}{y_{t-1}^2 x_{t-1}} \right) \right] - 1$$

(20)

which depends on $\sigma_u^2$ through $E\left( \frac{u_{t-1}^2}{y_{t-1}^2 x_{t-1}} \right)$ and it has no closed form expression. The bias of the ML estimator of $\mu$ is therefore

$$E[\hat{\phi}_T] - \phi = (e^\mu - 1 - \phi) + e^\mu E\left( \frac{u_{t-1}^2}{y_{t-1}^2 x_{t-1}} \right)$$

and it is composed of two parts: the first is due to the crude discretization, the second $e^\mu E\left( \frac{u_{t-1}^2}{y_{t-1}^2 x_{t-1}} \right)$ is the contribution of the measurement error. Even when $\sigma_u^2 = 0$ there is no closed form expression for the indirect inference estimators of $\mu$ and $\sigma$ in (17). Therefore, we look at the finite sample bias of the indirect inference estimates, $\hat{\mu}_{ST}$ and $\hat{\sigma}_{ST}$, by means of Monte Carlo simulations. We simulate 2000 observations from the gBm with $\mu_0 = 0.0004$, $\sigma_0 = 0.012$. We assume that $u_t \sim i.i.d.N(0, \sigma_u^0)$, where $\sigma_u$ in the range between 0 and 0.02. The discretization step is set to $\Delta = 1/24$. Figure 1 reports the percentage relative bias of the indirect inference estimator with respect to different values of $\sigma_u$. When the measurement error is absent, i.e. $\sigma_u = 0$, the binding function of $\hat{\phi}_T$ is $e^\mu - 1$ and the indirect inference gets rid of the discretization error, so that the estimates of $\mu$ and $\sigma$ are unbiased, see Gouriéroux and Monfort (1996). The impact of the measurement error on the bias remains limited, i.e. below 5%, when $\sigma_u$ is smaller than 0.003. As $\sigma_u$ increases, the percentage bias increases non-linearly for both $\hat{\mu}_{ST}$ and $\hat{\sigma}_{ST}$. When $\sigma_u$ is 0.02, i.e. almost two times as large as $\sigma_0$, then the percentage bias of $\hat{\sigma}_{ST}$ is more than 100% and that of $\hat{\mu}_{ST}$ is around 70%. Concluding, the zero-mean measurement error impacts not only on the estimate of the diffusion parameter $\sigma$, but also on the estimate of $\mu$. This is the result of the nonlinear transformation of the data used in the estimation, i.e. $w_t = x_t/x_{t-1}$, inducing the moment of a non linear transformation of the measurement error to appear in the binding function, as in (20).

2.2.2 Ornstein-Uhlenbeck process

An Ornstein-Uhlenbeck process is the solution of the differential equation:

$$dz(t) = k(\omega - z(t))dt + \sigma dW(t), \quad t > 0$$

(21)
consequence of the fact that the error term is an additive i.i.d process with zero mean, so that \( \omega \) is the mean parameter. When measurement error is absent, then the plim of \( \hat{\mu} \) and \( \hat{\sigma} \) of \( \hat{\mu} \) and \( \hat{\sigma} \) is the discretized process in (21), where the discretization step is \( \Delta = 1 \), so that \( x_t = x_{t-1} / \Delta = 1 + \beta_1 (\beta_2 - x_{t-1}) + \beta_3 e_t, \ e_t \sim i.i.d.N(0,1) \) (22) It follows that the set of auxiliary parameters is \( \beta = (\beta_1, \beta_2, \beta_3)' \). Let \( \theta_0 = (k_0, \omega_0, \sigma_0)' \) be the set of unknown true model parameters. The asymptotic bias of the indirect inference estimator \( \hat{\theta}_{ST} \) can be derived noting that the term \( p \lim (a^*/\sqrt{T}) = \left( \partial b(\theta_0)/\partial \theta \right)^{-1} [b(\theta_0, \psi_0) - b(\theta_0)] \) in equation (6), is equal to

\[
P \lim \left( \frac{a^*}{\sqrt{T}} \right) = \left[ \begin{array}{ccc} e^{-k_0} & 0 & 0 \\ 0 & 1 & 0 \\ e^{-k_0} & 1 & 0 \end{array} \right] \left( \begin{array}{c} \frac{2k_0}{1-e^{-2k_0}} \\ \frac{e^{-2k_0}}{2k_0} - \frac{1-e^{-2k_0}}{2k_0} \\ 0 \end{array} \right)^{1/2} \times \left( \begin{array}{c} e^{-k_0} - \frac{e^{-k_0} \sigma^2}{\sigma^2 + 2k_0 \sigma^2_{u,t}} \\ 0 \end{array} \right), \right.

(23)

\[
\left[ \left( \frac{\sigma^2}{2k_0} + \sigma^2_{u,t} \right) \left[ 1 + \left( \frac{e^{-k_0} \sigma^2}{\sigma^2 + 2k_0 \sigma^2_{u,t}} \right)^2 \right] - \left( \frac{e^{-k_0} \sigma^2}{\sigma^2 + 2k_0 \sigma^2_{u,t}} \right) e^{-k_0} \sigma^2_{u,t} \right]^{1/2} \right.

(24)

When \( \sigma^2_{u,t} > 0 \), then the plim of \( a^*/\sqrt{T} \) in the expansion of \( \hat{\theta}_{ST} \) is non null. Therefore, the indirect inference estimates of \( k \) and \( \sigma \) are asymptotically biased. Interestingly, the estimator of the long-run mean parameter \( \omega \) is not affected by the measurement error also when \( \sigma^2_{u,t} > 0 \). This is a consequence of the fact that the error term is an additive i.i.d process with zero mean, so that \( x_t \) and \( y_t \) have the same long-run mean. Moreover, it is easy to see that, when \( \sigma^2_{u,t} = 0 \), i.e. the measurement error is absent, then the plim of \( a^*/\sqrt{T} \) is zero so that \( \hat{\theta}_{ST} \) is consistent.

\[1\] Details on the derivation of an expression of \( p \lim a^*/\sqrt{T} \) for the Ornstein-Uhlenbeck process are in Appendix A.1.
3 A consistent indirect inference estimator

So far we have seen the impact that the noise has on the indirect inference estimator. A simple solution to the inconsistency caused by the presence of measurement error is to consider the nuisance parameters $\psi$ among the structural parameters that need to be estimated. The $(p+h) \times 1$ parameter vector to be estimated is now denoted by $\zeta = (\theta', \psi')'$. The auxiliary model is characterized by a criterion function $Q_T(x_T, \beta)$, where $\beta \in B$ with $B$ compact subset of $\mathbb{R}^q$, with $q \geq p + h$.

The proposed indirect inference procedure requires that we can simulate trajectories from the structural model contaminated with the measurement error. This means that we simulate $y_t^{(s)}$ from the structural model and the measurement error from the assumed conditional density. The contaminated artificial series, i.e. $x_t^{(s)} = y_t^{(s)} + u_t^{(s)}$, are used in place of $y_t^{(s)}$, thus

$$\hat{\beta}_T^s(\zeta) = \arg \min_{\beta} Q_T(x_1^{(s)}(\zeta), \ldots, x_T^{(s)}(\zeta); \beta).$$

The estimated binding function, $\hat{b}(\zeta)$, which now explicitly depends on $\theta$ and $\psi$, is used to match both sets of parameters.

**Assumption 4** Furthermore, similarly to Gouriéroux et al. (1993), we assume that:

i. the normalized function $Q_T(x_T^{(s)}(\zeta), \beta)$ uniformly converges in $(\zeta, \beta)$ to a deterministic function $Q_\infty(\zeta, \beta)$ when $T$ diverges.

ii. The limit function $Q_\infty(\zeta, \beta)$ has a unique maximum with respect to $\beta$. The maximum is $b(\zeta) = \arg \max_{\beta \in B} Q_\infty(\zeta, \beta)$.

iii. The functions $Q_T(x_T^{(s)}(\zeta), \beta)$ and $Q_\infty$ are differentiable with respect to $\beta$.

iv. The only solution of the asymptotic first order condition is associated with $\beta_0 = b(\zeta_0)$.

v. $b(\zeta)$ is a one-to-one (locally injective) function and $\frac{\partial b(\zeta_0)}{\partial \zeta}$ is a full-column rank matrix.

Assumption 4.v guarantees that $b(\zeta)$ is locally identified, so that the equation $\beta = b(\zeta)$ admits a unique solution in $\zeta$ at the true parameter value, $\zeta_0$. The indirect estimator of $\zeta$, i.e. of the structural and nuisance parameters, is obtained as

$$\hat{\zeta}_{ST} = \arg \min_{\zeta} \left\{ \hat{\beta}_T - \frac{1}{S} \sum_{s=1}^S \hat{\beta}_T^s(\zeta) \right\}^2$$

(25)

**Proposition 3.1** Under Assumptions 1-4 the indirect inference estimator $\hat{\zeta}_{ST}$ is consistent. Moreover, under regularity conditions, for $T \to \infty$ and $S$ fixed, the indirect inference estimator $\hat{\zeta}_{ST}$ is asymptotically normal, with

$$\sqrt{T}(\hat{\zeta}_{ST} - \zeta_0) \xrightarrow{d} N\left(0, W(S, \Omega)\right)$$

(26)

where $W(S, \Omega)$ is given in Gouriéroux and Monfort (1996, p.70).

**Proof:** See Gouriéroux et al. (1993).

The proof Proposition 3.1 follows directly from the results in Gouriéroux et al. (1993). Indeed, Assumptions 1-4 guarantee that the indirect inference problem at hand is standard, so that the asymptotic distribution of $\hat{\zeta}_{ST}$ is the same as in Proposition 3 in Gouriéroux et al. (1993). In other words, indirect inference provides asymptotically unbiased and normal estimates in presence of measurement error, if the parameters governing the latter are considered among the structural ones and pseudo-data can be simulated from the contaminated structural model. If the auxiliary model is such that it identifies all structural parameters, i.e. condition 4.v is satisfied, then standard theory applies.
In the next Section, we analyze the identification condition for the case of ARMA models. We establish a necessary condition for the identification of the ARMA parameters and the variance of the measurement error, i.e. that Assumption 4.\textsuperscript{v} holds.

### 3.1 Identification of ARMA models

This section briefly discusses the identification condition for the indirect estimation of a stationary ARMA\((r,l)\) signal observed with additive measurement error:

\[
\begin{align*}
x_t &= y_t + u_t & t = 1, \ldots, T \\
\alpha(L)y_t &= c + \varphi(L)e_t & e_t \sim i.i.d. N(0, \sigma^2_e) \tag{27}
\end{align*}
\]

We assume that the roots of \(\alpha(L) = 1 - \alpha_1L - \ldots - \alpha_pL^p\) and \(\varphi(L) = 1 + \varphi_1L + \ldots + \varphi_qL^q\) all lie outside the unit circle, and there are no common roots. Note that

\[
\alpha(L)x_t = c + \varphi(L)e_t + \alpha(L)u_t \tag{28}
\]

is an ARMA\((r, \max\{r, l\})\). The parameter vector to be estimated is \(\zeta = (c, \alpha_1, \ldots, \alpha_r, \varphi_1, \ldots, \varphi_l, \sigma^2_e, \sigma^2_u)'\).

The auxiliary model is an AR\((m)\), i.e.

\[
\phi(L)x_t = \phi_0 + \epsilon_t, \tag{29}
\]

where \(E(\epsilon_t) = 0, E(\epsilon_t^2) = \sigma^2_e\) and \(\phi(L) = (1 - \phi_1L - \ldots - \phi_mL^m)\). The \((q \times 1)\) vector of auxiliary parameters is \(\beta = (\phi_0, \phi_1, \ldots, \phi_m, \sigma^2_e)',\) with \(q = (m + 2) \geq p + 1 = 3 + r + l\).

Lemma 2.5 in Chanda (1995) makes clear that the identifiability of \(\varphi_1, \ldots, \varphi_l, \sigma^2_e\) and \(\sigma^2_u\) is possible if and only if \(r > l\). In the following proposition we explicit the identification condition of \(\zeta_0\) (Assumption 4.\textsuperscript{v}) when the auxiliary model is an AR\((m)\).

**Proposition 3.2** Let the structural model be the stationary ARMA\((r, l)\) in (27) with measurement error \(u_t \sim WN(0, \sigma^2_u)\) and the auxiliary model be the AR\((m)\) in (29) with \(m > r + l\). The binding function is

\[
b(\zeta) = \begin{bmatrix} Q_{ZZ}^{-1}Q_{ZX} \\ Q_{XX} - Q_{XZ}Q_{ZZ}^{-1}Q_{ZX} \end{bmatrix}
\]

where \(Q_{ZZ}, Q_{XX}\) and \(Q_{ZX}\) contain the mean, the variance and the autocovariances of the process for \(x_t\) in (28) up to lag \(m\). Then the Jacobian matrix

\[
\frac{\partial b(\zeta)}{\partial \zeta'} = \begin{bmatrix} Q_{XX} & -(Q_{ZZ}^{-1} \otimes Q_{XX}^{-1}) \frac{\partial \vec{Q}_{XX}}{\partial \zeta'} + Q_{ZZ}^{-1} \frac{\partial Q_{XX}}{\partial \zeta'} \\ -\text{vec}(Q_{XX}Q_{XZ})'(Q_{ZZ}^{-1} \otimes Q_{ZZ}^{-1}) \frac{\partial \vec{Q}_{XX}}{\partial \zeta'} - 2Q_{XZ}Q_{ZZ}^{-1} \frac{\partial Q_{XX}}{\partial \zeta'} \end{bmatrix}
\]

has full column rank in \(\zeta_0\) if \(r > l\).

**Proof:** See Appendix B.2.

The proposition says that the ARMA structural model is not identified when the number of the moving average parameters is larger or equal to the number of autoregressive coefficients, even if \(q \geq p + 1\). In this case the Assumption 4.\textsuperscript{v} does not hold for any choice of the auxiliary AR\((m)\) model with \(m > l + r\). In other words, if the structural model is not identifiable, then there is no possibility also for the indirect inference estimator to guarantee consistency. Therefore, the condition \(r > l\) is a necessary condition for the identification of ARMA plus noise processes by AR\((m)\) models.\textsuperscript{2}

\textsuperscript{2}We conjecture that this condition is also sufficient. But the proof is fairly involved, since it would consist of showing that full-column rank of the Jacobian implies \(r > l\).
The case of an MA(1) contaminated by measurement error provides an example of the violation of the identification condition. Consider again the example in Section 2.1. In this case, the set of structural parameters is $\zeta = [\varphi, \sigma_u^2, \sigma_\varepsilon^2]^T$. Therefore, the number of parameters in the auxiliary model must be at least 3 to satisfy the order condition $q \geq p + 1$ for indirect inference. Suppose the auxiliary model is a zero-mean AR(2). In this case, the $(3 \times 1)$ vector of auxiliary parameters is $\beta = [\phi_1, \phi_2, \sigma_\varepsilon^2]^T$. The binding function is given by

$$b(\zeta) = p \lim_{T \to \infty} \begin{bmatrix} \hat{\phi}_{1,T} \\ \hat{\phi}_{2,T} \\ \hat{\sigma}_\varepsilon^2,T \end{bmatrix} = \begin{bmatrix} \frac{(1+\varphi^2)(\sigma_u^2+\sigma_\varepsilon^2)\omega^2}{(1+\varphi^2)(\sigma_u^2+\sigma_\varepsilon^2)\omega^2 - \varphi^2\sigma_\varepsilon^4} \\ \frac{-\varphi^2\sigma_\varepsilon^4}{(1+\varphi^2)(\sigma_u^2+\sigma_\varepsilon^2)\omega^2 - \varphi^2\sigma_\varepsilon^4} \\ (1+\varphi^2)\sigma_u^2 + \sigma_\varepsilon^2 + 3\frac{[(1+\varphi^2)(\sigma_u^2+\sigma_\varepsilon^2)\omega^2\sigma_\varepsilon^4]}{(1+\varphi^2)(\sigma_u^2+\sigma_\varepsilon^2)\omega^2 - \varphi^2\sigma_\varepsilon^4} \end{bmatrix}$$

and a closed form expression of $\frac{\partial b(\zeta)}{\partial \zeta}$ can be obtained. However, $\frac{\partial b(\zeta)}{\partial \zeta}$ has rank equal to 2 for any $\zeta$ in the parameter space, while the number of structural parameters is 3. This is a consequence of the fact that the ARMA condition $r > l$ is not satisfied for the MA(1) case, meaning that the structural model cannot be identified by any AR(m) with $m > 1$.

### 3.2 Estimation of the Ornstein-Uhlenbeck process with measurement error

As noted in Section 2.2.2, indirect inference corrects for the discretization error when data are generated by an Ornstein-Uhlenbeck process and a crude discretization scheme is applied. However, when the discretized data are observed with measurement error, and this is neglected, then a bias term emerges in the indirect inference estimates. When instead the variance of the measurement error is included among the structural parameters, the $(4 \times 1)$ vector of structural parameters is $\zeta = [k, \omega, \sigma, \sigma_u^2]^T$. The auxiliary model in equation (22), which contains only 3 parameters, must be extended to satisfy the order condition, $q \geq p + 1$. As additional parameter we choose, $\sigma_e^2 = \text{Var}(x_1)$, so that the vector of auxiliary parameters is $\beta = (\beta_1, \beta_2, \beta_3, \sigma_e^2)^T$. It is straightforward to extend the results of Section 2.2.2 to this case, so that the binding function is

$$b(\zeta) = \left[ \frac{1 - e^{-k\sigma_u^2}}{\sigma_u^2 + 2k\sigma_u^2} \right] \omega \left[ \left( \frac{\sigma_u^2 + \sigma_e^2}{\sigma_u^2 + 2k\sigma_u^2} \right) \right] - \left( \frac{e^{-k\sigma_u^2}}{\sigma_u^2 + 2k\sigma_u^2} \right) \left( \frac{e^{-k\sigma_u^2}}{\sigma_u^2 + 2k\sigma_u^2} \right) \right]^{1/2}$$

and the Jacobian matrix $\frac{\partial b(\zeta)}{\partial \zeta}$ has full rank. Hence, the auxiliary model identifies all the parameters in $\zeta$, including the variance of the measurement error.

In Table 1 we report the bias and RMSE of the indirect estimates of the Ornstein-Uhlenbeck parameters based on Monte Carlo simulations. In each Monte Carlo simulation, an i.i.d. Gaussian noise is added to the discretized trajectories of the Ornstein-Uhlenbeck process, so that indirect inference is carried out on the contaminated series. The bias and RMSE of the indirect estimates that consider the presence of the measurement error as part of the vector of structural parameters are reported in the second and third columns. It emerges that the indirect estimates of $k$ are almost unbiased in all cases considered, i.e. for $k = 0.8$ or 0.05 and for the two values of the noise-to-signal ratio, defined as $\text{nsr} = \frac{\sigma_e^2}{\text{Var}(y_t)}$. The same can be said for the diffusion coefficient ($\sigma$) whose estimate is centered on the true value. Moreover, the finite-sample bias of the estimates of $\sigma_u^2$ is small for both $\text{nsr}$ values.

---

3Details on the derivation of these results are shown in Appendix A.2. Due to space constraints the expression of $\frac{\partial b(\zeta)}{\partial \zeta}$ is not reported. It is available upon request from the authors.
<table>
<thead>
<tr>
<th>nsr=0.3</th>
<th>(\zeta)</th>
<th>(\theta)</th>
<th>Signal</th>
</tr>
</thead>
<tbody>
<tr>
<td>(k)</td>
<td>0.8</td>
<td>0.8275</td>
<td>0.0568</td>
</tr>
<tr>
<td>(\omega)</td>
<td>0.1</td>
<td>0.1001</td>
<td>0.0117</td>
</tr>
<tr>
<td>(\sigma)</td>
<td>0.06</td>
<td>0.0618</td>
<td>0.0186</td>
</tr>
<tr>
<td>(\sigma_u)</td>
<td>0.026</td>
<td>0.0245</td>
<td>0.0034</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>nsr=0.5</th>
<th>(\zeta)</th>
<th>(\theta)</th>
<th>Signal</th>
</tr>
</thead>
<tbody>
<tr>
<td>(k)</td>
<td>0.8</td>
<td>0.8409</td>
<td>0.0551</td>
</tr>
<tr>
<td>(\omega)</td>
<td>0.1</td>
<td>0.1001</td>
<td>0.0117</td>
</tr>
<tr>
<td>(\sigma)</td>
<td>0.06</td>
<td>0.0628</td>
<td>0.0232</td>
</tr>
<tr>
<td>(\sigma_u)</td>
<td>0.0335</td>
<td>0.0319</td>
<td>0.0035</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>nsr=0.3</th>
<th>(\zeta)</th>
<th>(\theta)</th>
<th>Signal</th>
</tr>
</thead>
<tbody>
<tr>
<td>(k)</td>
<td>0.05</td>
<td>0.0622</td>
<td>0.0233</td>
</tr>
<tr>
<td>(\omega)</td>
<td>0.1</td>
<td>0.1003</td>
<td>0.0081</td>
</tr>
<tr>
<td>(\sigma)</td>
<td>0.02</td>
<td>0.0223</td>
<td>0.0041</td>
</tr>
<tr>
<td>(\sigma_u)</td>
<td>0.0346</td>
<td>0.0338</td>
<td>0.0023</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>nsr=0.5</th>
<th>(\zeta)</th>
<th>(\theta)</th>
<th>Signal</th>
</tr>
</thead>
<tbody>
<tr>
<td>(k)</td>
<td>0.05</td>
<td>0.0634</td>
<td>0.0284</td>
</tr>
<tr>
<td>(\omega)</td>
<td>0.1</td>
<td>0.1003</td>
<td>0.0082</td>
</tr>
<tr>
<td>(\sigma)</td>
<td>0.02</td>
<td>0.0224</td>
<td>0.0050</td>
</tr>
<tr>
<td>(\sigma_u)</td>
<td>0.0447</td>
<td>0.0440</td>
<td>0.0023</td>
</tr>
</tbody>
</table>

Table 1: Ornstein-Uhlenbeck process. Mean and RMSE of estimated parameters with and without the inclusion of the variance of the measurement error among the structural parameters. Last two columns report the parameter estimates based on an observed OU process generated without noise (Signal). nsr stands for noise-to-signal ratio, \(nsr = \text{Var}(u_t) / \text{Var}(y_t)\). The simulation are carried out with Euler discretization scheme with step size corresponding to 1 day. The number of simulated days is \(T = 1000\), and the number of simulated points is \(M = 1000\).

In the last two columns, Table 1 reports the mean and the RMSE of the unfeasible estimates, i.e. those based on the non-contaminated discretized trajectory, \(y_t\). These estimates constitute a natural benchmark to quantify the efficiency loss that arises when the indirect inference is based on \(x_t\) and the variance of the measurement error is estimated. As expected, the bias of the unfeasible estimator is almost null and the uncertainty of the estimates, as measured by the RMSE, is the lowest for all parameters. The main loss in efficiency of the feasible estimator of \(\zeta\) is related to the volatility parameter \(\sigma\), as the RMSE is 5 times larger than in the unfeasible case. This is not surprising as the feasible indirect inference estimator must be able to disentangle the total observed variance into the variance of the latent innovation and variance of the measurement error, and this produces more uncertainty in the estimates. The RMSE of the feasible estimator is however much lower than that achieved by the indirect inference estimator that neglects the noise (fourth and fifth column), with the exception of the estimates of \(\omega\) that are not affected by the noise, as proved in equation (23). In particular, the estimates of \(k\) and \(\sigma\) are dramatically affected by the presence of measurement error when the variance of the latter is not included among the structural parameters.
4 Estimation of SV models with RV

The recent results in the theory of RV based on high frequency data open the door to the estimation of continuous-time SV models by GMM and indirect inference. Under unrealistic assumptions, e.g. the absence of microstructure noise (MN), the RV is an asymptotically unbiased and efficient estimator of IV. However, the presence of MN can dramatically affect the consistency of the SV parameter estimates. The MN is generated by structural features of financial markets, like trading rules, the bid-ask spread and the discreteness of price changes. Neglecting the MN in calculating RV leads to biased and inconsistent estimates of integrated volatility (IV) as a true measure of daily volatility. Indeed, when the sample interval shrinks to zero, the MN obscures the IV signal.

In a GMM framework, Bollerslev and Zhou (2002) propose a simplified approach to deal with this problem which does not disentangle the discretization error from the MN. Differently, Corsi and Renò (2012), in an indirect inference framework, sample log-returns at low frequencies and do not explicitly model the MN. We call this approach neutralization.

The problem of the estimation of the parameters of SV models by RV under the presence of MN has been tackled also in Corradi and Distaso (2006) and Todorov (2009). Corradi and Distaso (2006) propose a SMM approach and derive a set of sufficient conditions for the asymptotic negligibility of the measurement error, when the moments of the unobservable IV are replaced by the moments of the RV. Alternatively, Todorov (2009) uses a corrected estimator robust to MN and price jumps, but he also derives the moment conditions in closed form so that there is no need for simulation. However, the use of corrected realized estimators, like those proposed in Zhang et al., 2005, Barndorff-Nielsen et al., 2008, Hansen et al., 2008 and Andersen et al., 2012, may not be the best solution in the indirect inference framework. Indeed, these realized estimators depend in a non trivial way on the nuisance parameters of the MN distribution, and their consistency is derived under some (strong) assumptions about the MN. In this case, the binding function would still depend on the MN nuisance parameters, making impossible to be matched by the simulated trajectories and hence potentially causing inconsistent estimates of the SV parameters.

Instead, consistently with the general approach outlined in this paper, we estimate the parameters of the market MN distribution along with the SV structural parameters. In an indirect inference setup, we contaminate the simulated trajectories from the SV model, avoiding in this way the need to neutralize the impact of the MN on the volatility estimates. Hence, the auxiliary model can be based on potentially distorted but efficient and simple estimators of IV, like RV and BPV. This represents a very general approach to the treatment of the problem of measurement error in the estimation of the SV framework, as it is potentially valid for any SV model and contamination scheme. Since the parameters of the MN distribution must be estimated jointly with the structural ones, special attention has to be devoted to the identification issue. In the following Section, we closely look at the identification of the Heston SV model parameters with MN.

4.1 Estimation of the Heston model with microstructure noise

The Heston (1993) model is a well known continuous-time stochastic process used to describe the evolution of the volatility of an underlying asset and widely used in option pricing. Assume that \( \sigma^2(t) \) follows a square root process as in Heston (1993), then

\[
\begin{align*}
    dp^*(t) &= \sigma(t)dW_1(t) \\
    d\sigma^2(t) &= \kappa(\omega - \sigma^2(t))dt + \varsigma \sigma(t)dW_2(t)
\end{align*}
\]

where \( \kappa > 0 \) governs the speed of mean reversion, \( \varsigma > 0 \) is the volatility of volatility parameter, while \( \omega > 0 \) is the long run mean of \( \sigma^2(t) \), where the latter is the instantaneous volatility and it is independent of the process \( W_1(t) \).\(^4\) The condition \( 2\kappa \omega \geq \varsigma^2 \) guarantees that the volatility process

\(^4\)For ease of exposition the assumption that \( \text{Corr}[dW_1(t), dW_2(t)] = 0 \), i.e. absence of leverage, is maintained. In the empirical analysis, this restriction is removed.
where \( f_t \) is a state-variable process. The function \( P_1(\cdot) \) is defined so that it has the following properties:

\[
E[P_1(f_t)] = 0, \\
\text{Var}[P_1(f(t))] = 1, \\
E[P_1(f(t + \Delta))|P_1(f(\tau)), \tau \leq t] = e^{-\lambda_1 \Delta} P_1(f(t)).
\]

Following Meddahi (2003), model (32) for \( \sigma^2(t) \) can be rewritten as in (33) with \( P_1(f(t)) = \frac{\sqrt{\omega}}{\sqrt{\omega_1}}(\omega - f(t)), a_0 = \omega, a_1 = -\varsigma \frac{\sqrt{\omega_1}}{\omega_1}, \lambda_1 = \kappa \) and \( \sigma^2(t) = f(t) \). Now, we focus on the properties of the ex-post estimates of IV, defined as \( IV = \int_{t-1}^t \sigma^2(u)du \), which cumulates the instantaneous volatility over periods of unit length. A non-parametric estimator of \( IV \) is \( RV = \sum_{i=1}^n \sigma^2_i(\Delta) \), where \( n = 1/\Delta \), and \( \sigma_i \) are the intraday returns over the intervals \( [t-1+i\Delta, t-1+i\Delta] \), for \( i = 1, \ldots, n \). When the MN is present and contaminates the high-frequency returns the observed intraday price is observed with error, i.e.

\[
p_{t,i}(\Delta) = p_{t,i}^*(\Delta) + \epsilon_{t,i}(\Delta) \quad \text{for and} \quad t = 1, \ldots, T \quad i = 1, \ldots, n
\]

where \( p_{t,i}^*(\Delta) \) is the \( i \)-th latent efficient log-price on day \( t \). The term \( \epsilon_{t,i}(\Delta) \) is the noise around the true price, with mean 0 and finite fourth moment and it is assumed i.i.d. and independent of the efficient price. Over periods of length \( \Delta \), the log-return \( r_{t,i}(\Delta) \equiv \log(1+\frac{\epsilon_{t,i}(\Delta)}{\sigma_{t,i}(\Delta)}) \) is given by

\[
r_{t,i}(\Delta) = (p_{t,i}(\Delta) - p_{t-1,i}(\Delta)) + (\epsilon_{t,i}(\Delta) - \epsilon_{t-1,i}(\Delta)) = r_{t,i}(\Delta) + \nu_{t,i}(\Delta)
\]

with \( \sigma^2_{\nu_i} = \text{Var}[\nu_{t,i}(\Delta)] \).

When there is no drift in prices, the \( RV \) is observed with a measurement error, that is due both to the discretization error and to the MN:

\[
RV(\Delta) = IV + \nu_{t,\Delta},
\]

where

\[
u_{t,\Delta} = \eta(\Delta) + \sum_{i=1}^n \nu_{t,i}^2(\Delta) + 2 \left( \sum_{j=1}^n \sigma_{t,i,\Delta} z_{t,\nu_{t,i,\Delta}} \right).
\]

\( \eta(\Delta) = \sum_{i=1}^n \eta_{t+\Delta} \cdot \Delta \) is the discretization error. Meddahi (2002) proves that \( \eta(\Delta) \) has a nonzero mean, when the drift in prices is non-zero, and is heteroskedastic. The correlation between IV and \( \eta(\Delta) \) is zero when there is no leverage effect (Barndorff-Nielsen and Shephard (2002b) and Meddahi (2002)).

Assuming that the drift is null and there is no leverage effect, then Barndorff-Nielsen and Shephard (2002a) show that, for finite \( \Delta > 0 \), the discretization error for the interval \( [t-1+i\Delta, t-1+i\Delta] \) can be written as

\[
\eta_{t-1+i\Delta}(\Delta) = \sigma^2_{t,i}(\Delta) \left( z_{t,i}^2 - 1 \right)
\]

where \( z_{t,i} \) is \( i.i.d. N(0,1) \) and it is independent of \( \sigma^2_{t,i}(\Delta) = \int_{t-1+i\Delta}^{t-\Delta} \sigma^2(s)ds \), \( \sigma^2_{t,i}(\Delta) \) is the integrated variance over the \( i \)-th subinterval of length \( \Delta \). Meddahi (2003) proves that when the instantaneous volatility is a square-root process, like in the Heston (1993) model, then both IV and RV have an ARMA(1,1) representation. As noted by Meddahi (2003), already Bollerslev and Zhou (2002) explicitly recognized that IV and RV are ARMA\((p,p)\) processes, \( p = 1, 2 \), when the spot variance depends on \( p \) square-root processes.
representations of IV and RV coincide. However, when the MN is present, then the variance of the measurement error, \( \text{Var}[u_t(\Delta)] \), does not converge to zero as \( \Delta \to 0 \), but it diverges thus completely obscuring the volatility signal. As noted by Bandi and Russel (2006), while the efficient return is of order \( O_p(\sqrt{\Delta}) \), the MN is of order \( O_p(1) \) over any period of time. This means, that, when \( \Delta \to 0 \), the MN dominates over the true return process, and longer period returns are less contaminated by noise than shorter period returns.

For a given \( \Delta > 0 \), the mean and the variance of RV are equal to
\[
E[RV_t(\Delta)] = E[IV_t] + E[u_t(\Delta)] \\
\text{Var}[RV_t(\Delta)] = \text{Var}[IV_t] + \text{Var}[u_t(\Delta)]
\]
with \( E[IV_t] = \omega, E[u_t(\Delta)] = \Delta^{-1}\sigma_u^2 \), and \( \text{Var}[u_t(\Delta)] = 2\Delta^{-1}E\left[\left(\sigma_u^2 t_i(\Delta)\right)^2\right](1 + 2\sigma_u^2) + \Delta^{-1}(\kappa_u - \sigma_u^2) \), \( \kappa_u = E[u_t(\Delta)^4] \). It follows that the variance of \( RV_t(\Delta) \) is
\[
\gamma(0) = \text{Var}[RV_t(\Delta)] = 2\frac{\sigma_u^2}{\kappa_u^2}[\exp(-\kappa) + \kappa - 1] + \text{Var}[u_t(\Delta)]
\]
and the autocovariances of \( RV_t(\Delta) \) are
\[
\gamma(j) = \text{Cov}[RV_t(\Delta), RV_{t-j}(\Delta)] = \text{Cov}[IV_t, IV_{t-j}] \\
= \frac{a_t^2}{\kappa_u^2}[1 - \exp(-\kappa)^2 \exp(-\kappa(j-1))] \quad j > 0.
\]

We are interested in estimating the parameters of the Heston model in (31)-(32) along with those of conditional distribution of the MN which contaminates the log-returns at high frequency. Estimates of the structural parameter vector, \( \zeta = (\kappa, \omega, \varsigma, \sigma_u^2)' \), can be obtained by indirect inference using an auxiliary model based on RV. A well known example of a simple reduced-form model for RV is the HAR-RV model of Corsi (2009), which is
\[
x_t = \phi_1 + \phi_2 x_{t-1} + \phi_3 x_{t-1}^w + \phi_4 x_{t-1}^m + e_t,
\]
where \( x_t = RV_t(\Delta), x_{t}^w = \frac{1}{\Delta} \sum_{i=0}^{1} x_{t-i} \) and \( x_{t}^m = \frac{1}{\Delta} \sum_{i=0}^{21} x_{t-i} \), and \( \phi = [\phi_1, \phi_2, \phi_3, \phi_4]' \). The (5 \times 1) vector of auxiliary parameters is \( \beta = (\phi', \sigma_u^2)^' \).

**Proposition 4.1 (Identification of Heston model with MN)** Let the structural model be the Heston model in (31) and (32), with the RV and the measurement error as in (36) and (37), respectively. The auxiliary model is the HAR-RV in (42), which is an AR(22) with restrictions contained in the (23 \times 4) matrix, \( R \). The binding function results to be
\[
b(\zeta) = \begin{bmatrix} R'Q_{ZZ}R^{-1} R'Q_{ZX} \\ Q_{XX} - [Q_{XZ} R (R'Q_{ZZ} R)^{-1} R'Q_{ZX}] \end{bmatrix}
\]
where \( Q_{ZZ}, Q_{XX}, Q_{XZ} \) are the moment matrices of RV and are function of the Heston parameters. The Jacobian matrix \( \frac{\partial b(\zeta)}{\partial \zeta} \) has full column rank in \( \zeta_0 \).

**Proof:** See Appendix B.3.

Proposition 4.1 proves that the Heston parameters can be identified when the auxiliary model is the HAR-RV since the rank of the binding function is full. At a first sight, this result seems to contradict the evidence reported in Section 3.1 about the non-identifiability of an ARMA(1,1) plus noise when \( l \leq r \). On the contrary, despite RV is the sum of an ARMA(1,1) and a noise
term, identification is guaranteed in this case. Indeed, as noted by Barndorff-Nielsen and Shephard (2002b) and Meddahi (2003), the moving-average root of the IV signal is in turn the result of the state-space representation of an AR(1) signal plus noise, so that the MA parameter is restricted and dependent on the autoregressive root. The consequence is that the AR and MA parameters are no more functionally independent, which means that the parameter space dimension is reduced by one. This explains why the ARMA(1,1) plus noise representation of RV is identifiable, in terms of the underlying Heston parameters, by an autoregressive auxiliary model.

4.1.1 Monte Carlo simulations

In the Monte Carlo experiments reported below we study the indirect inference estimation of the continuous-time SV model in (31) and (32) using the series of non-overlapping RV’s, i.e. \{RV_i(\Delta)\}_{i=1}^T under the presence of MN. Tables 2-3 report the Monte Carlo summary statistics of the indirect inference estimates of the Heston model parameters with and without accounting for the presence of MN, using different sampling frequencies, 30 seconds, 1, 5 and 30 minutes to construct the daily RV.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(\theta)</td>
<td>(\zeta)</td>
</tr>
<tr>
<td></td>
<td>(\kappa)</td>
<td>(\omega)</td>
</tr>
<tr>
<td>30 sec</td>
<td>0.1016</td>
<td>0.8889</td>
</tr>
<tr>
<td>1 min</td>
<td>0.1014</td>
<td>0.6925</td>
</tr>
<tr>
<td>5 min</td>
<td>0.1010</td>
<td>0.5346</td>
</tr>
<tr>
<td>30 min</td>
<td>0.0965</td>
<td>0.4911</td>
</tr>
</tbody>
</table>

Table 2: Heston SV model. Mean and RMSE of estimated parameters with and without correction for the MN. The true parameter set is \(\kappa = 0.1\), \(\omega = 0.5\), \(\varsigma = 0.1\) and \(\sigma^2_{\nu} = 0.0005\). The simulation are carried out with Euler discretization scheme with step size corresponding to 30 seconds. Four aggregation levels are considered to construct RV, 30 seconds and 1, 5 and 30 minutes. The number of simulated days is \(T = 1500\) with \(n = 780\) intradaily observations are generated in each day. The number of Monte Carlo simulations is \(M=1000\).

The true parameter set is calibrated to values close to those found in empirical works, see Bollerslev and Zhou (2002) and Garcia et al. (2011), and are relative to percentage returns. The long-run mean parameter, \(\omega\), is set equal to 0.5 and 0.8, which corresponds to an annualized volatility of 11.2% and 14.2% respectively. The speed of mean reversion, \(\kappa\), is equal to either 0.1 or 0.05, while the vol-of-vol parameter \(\varsigma\) is either 0.1 or 0.2. For each Monte Carlo replication, a simulated trajectory for the intradaily returns is generated from Euler discretization of model (32) for \(T = 1500\) days, with intradaily step of 30 seconds, which corresponds to \(n = 780\). The intraday return is contaminated with an additive MN, \(\nu_{i,j}\), that is assumed to be i.i.d. Gaussian with mean 0 and variance \(\sigma^2_{\nu} = 0.0005\). This choice for \(\sigma^2_{\nu}\) is in line with the numbers reported in Aït-Sahalia

\(^7\)As in Bandi and Russell (2006), we contaminate the intradaily log-returns instead of the intradaily log-prices.
et al. (2005, p.364), and corresponds to a percentage standard deviation of 0.02%. According to Aït-Sahalia et al. (2005), the optimal sampling frequency, under the assumption of constant volatility, should be between 1 and 5 minutes. The RV series is constructed from high-frequency returns with different sub-sampling and, similarly to Corsi and Renò (2012), the HAR-RV model on log$(RV_t)$ is used as auxiliary model such that the set of auxiliary parameters in the indirect inference estimation is $\beta = [\phi_1, \phi_2, \phi_3, \phi_4, \sigma_\epsilon^2]^T$. The approximation of the binding function is based on $S = 100$ simulated trajectories of equation (32) with the same Euler discretization as the sampling frequency used to compute the RV series on the real data. When the possible presence of MN is neglected in the indirect inference estimation, the set of structural parameters is $\theta = [\kappa, \omega, \varsigma]^T$. On the contrary, if the MN is considered in the indirect inference estimation, the set of structural parameters, $\zeta$, also includes $\sigma_\nu^2$. In this case, the simulated trajectories of the log-returns used in the second stage of the indirect estimation are contaminated with a Gaussian MN with variance $\sigma_\nu^2$. In both cases, the HAR-RV model is used as auxiliary model on each simulated series, $RV_t^{(s)}$, for $s = 1, ..., S$, and the set of parameters $\hat{\beta}_s$ is estimated by OLS.

<table>
<thead>
<tr>
<th>Mean</th>
<th>[\kappa]</th>
<th>[\omega]</th>
<th>[\varsigma]</th>
</tr>
</thead>
<tbody>
<tr>
<td>30 sec</td>
<td>0.0493</td>
<td>1.2008</td>
<td>0.1514</td>
</tr>
<tr>
<td>1 min</td>
<td>0.0482</td>
<td>1.0027</td>
<td>0.1654</td>
</tr>
<tr>
<td>5 min</td>
<td>0.0506</td>
<td>0.8448</td>
<td>0.1892</td>
</tr>
<tr>
<td>30 min</td>
<td>0.0514</td>
<td>0.8086</td>
<td>0.1969</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>RMSE</th>
<th>[\kappa]</th>
<th>[\omega]</th>
<th>[\varsigma]</th>
</tr>
</thead>
<tbody>
<tr>
<td>30 sec</td>
<td>0.0153</td>
<td>0.4159</td>
<td>0.0542</td>
</tr>
<tr>
<td>1 min</td>
<td>0.0144</td>
<td>0.2338</td>
<td>0.0424</td>
</tr>
<tr>
<td>5 min</td>
<td>0.0131</td>
<td>0.1239</td>
<td>0.0267</td>
</tr>
<tr>
<td>30 min</td>
<td>0.0163</td>
<td>0.1223</td>
<td>0.0312</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>RMSE</th>
<th>[\sigma_\nu^2]</th>
</tr>
</thead>
<tbody>
<tr>
<td>30 sec</td>
<td>0.0161</td>
</tr>
<tr>
<td>1 min</td>
<td>0.0123</td>
</tr>
<tr>
<td>5 min</td>
<td>0.0284</td>
</tr>
<tr>
<td>30 min</td>
<td>0.0005</td>
</tr>
</tbody>
</table>

Table 3: Heston $SV$ model. Mean and RMSE of estimated parameters with and without correction for the MN. The true parameter set is $\kappa = 0.05$, $\omega = 0.8$, $\varsigma = 0.2$ and $\sigma_\nu^2 = 0.0005$. The simulation are carried out with Euler discretization scheme with step size corresponding to 30 seconds. Three aggregation levels are considered to construct RV, 30 seconds and 1, 5 and 30 minutes. The number of simulated days is $T = 1500$ with $n = 780$ intradaily observations generated in each day. The number of Monte Carlo simulations is $M=1000$.

The results presented in Tables 2-3, based on $M = 1000$ Monte Carlo simulations, reflect the properties of RV under MN, see Aït-Sahalia et al. (2005). If we don’t want to take into account the MN, the best choice, in terms of bias, is to compute the RV using a low sampling frequency, say 5 or 30 minutes. Indeed, sampling at 30 minutes neutralizes the effect of the MN on the estimates of the RV, such that the indirect inference estimates are only slightly affected. For example, the bias of $\hat{\omega}_{30\text{min}}$ and $\hat{\varsigma}_{30\text{min}}$ are negligible in all cases. On the other hand, sampling at 30 seconds or at 1 minute, but neglecting the presence of MN, induces large biases in the indirect inference estimates of $\omega$, which as expected is over-estimated, and in $\varsigma$. Surprisingly, the estimates of $\kappa$ are unbiased for any choice of $\Delta$ also when the MN is neglected in the estimation. In terms of efficiency, the estimates corresponding to a 5-minutes sampling scheme without correction are

\[\text{This leads to equivalent results as } \sigma_\nu^2 = 2\sigma_\epsilon^2, \text{ where } \sigma_\epsilon^2 \text{ is the variance of } \epsilon_{t,i} \text{ in } p_{t,i} = p_{t,i}^* + \epsilon_{t,i}.\]

19
generally associated with a lower RMSE than those obtained sampling at 30 minutes. Indeed, the effect of the MN when sampling returns every 5 minutes is rather limited, and the efficiency of RV when n is relatively large leads to more precise estimates of the Heston parameters than those obtained when the sampling frequency is 30 minutes. In other words, the squared bias component of RMSE when sampling at 5 minutes is smaller than the variance component of RMSE when sampling at 30 minutes. Moreover, when sampling at 30 seconds, the squared bias term dominates over the parameter variance in the RMSE, especially for ω and δ.

Turning our attention to the results when the variance of the MN is included among the structural parameters, it emerges that sampling at the highest possible frequency provides unbiased estimates of all structural parameters. Instead, the estimates of σ^2 are biased when sampling at low frequencies, since the log-returns do not contain enough information about the MN. Therefore, if one is interested in estimating the micro-structure noise together with the SV parameters, the best option is to sample at the highest frequency available. Notably, in Table 2, the RMSE of the estimates of ω and δ obtained sampling at 30 seconds intervals is significantly smaller than the RMSE obtained by neutralizing the microstructure noise, i.e. sampling at 5 or 30 minutes. This means that the impact of the discretization error, that is a function of the Heston parameters via the term $2\Delta^{-1}E\left(\sigma^2_{t,i}(\Delta)\right)^2$ in the variance of $u_t$, affects more significantly the SV parameter estimates when the latent volatility process is not very persistent and with a smaller long-run mean, as in Table 2. In these cases, the price paid by sampling log-returns at low frequencies is relatively higher than in the setup characterized by a persistent instantaneous volatility. Indeed, when the volatility signal is more persistent, with larger long-run mean and vol-of-vol, as in Table 3, the values of the RMSE at 30 seconds are roughly the same as those obtained neutralizing the MN and sampling at 5 minutes. This means that the impact of the discretization error is less relevant in this case, and that sampling at low frequency does not induce severe efficiency losses.

5 Empirical application

We provide some empirical evidence of the importance of accounting for the MN, when performing indirect inference estimates of SV based on RV series. In particular, we estimate the parameters of the two-factor Heston model (TFSV henceforth) on the RV series of JPMorgan (JPM) from July 2, 2003 to June 29, 2007 using intradaily returns sampled at 5-seconds frequency, i.e. n = 4680. The choice of the sample period is motivated by the evidence of parameter instability for the TFSV model during the sub-prime crisis, between June-2007 until June 2009, as shown in Grassi and Santucci de Magistris (2015). Instead, in the period 2003-2007, the RV is not subject to major breaks and we expect the parameters to be rather stable through time. The RV series is computed with returns sampled at two frequencies, 5-seconds and 5-minutes, RV^5s and RV^5m respectively. The dynamics of the two series are reported in Figure 2. The impact of the MN emerges clearly from the graph since the long run mean of RV^5s is shifted upward compared to that of RV^5m. Moreover, RV^5m is more noisy than RV^5s, meaning that the discretization error, denoted by η(Δ) in equation (37) seems to have a higher impact on the variance of the measurement error while the MN mainly impacts on the mean of RV. This evidence is also confirmed by the sample statistics reported in Table 4. The moments of the daily de-volatized returns, $\tilde{r}_t = r_t/\sqrt{RV^5m_t}$, are rather close to those of the standard Gaussian distribution. Notably, the autocorrelation function of RV^5s is much higher than that of RV^5m, as a consequence of the smaller impact of the discretization error on the variance of the measurement error.

The BPV computed from returns sampled at 5-seconds is clearly downward biased, as its mean and variance are much lower than those of RV^5m. Conversely, the moments of RV^5m are very close to those of BPV^5m. The reason for the bias in BPV^5s is the decimalization effect, which induces discontinuities in the trajectories of returns sampled at very high-frequencies, so that the product $|r_{t,i}| \times |r_{t,i-1}|$ is often equal to 0.


Figure 2: RV of JPM based on 5 seconds (red) and 5 minutes (black) sampling.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>SK</th>
<th>KU</th>
<th>AR(1)</th>
<th>AR(20)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tilde{r}_t$</td>
<td>0.0516</td>
<td>0.9019</td>
<td>0.0572</td>
<td>2.8371</td>
<td>-0.0074</td>
<td>0.0030</td>
</tr>
<tr>
<td>$RV_{5s}$</td>
<td>1.4073</td>
<td>0.6656</td>
<td>2.1826</td>
<td>11.454</td>
<td>0.6423</td>
<td>0.2861</td>
</tr>
<tr>
<td>$RV_{5m}$</td>
<td>1.0266</td>
<td>0.7381</td>
<td>2.9010</td>
<td>16.415</td>
<td>0.5204</td>
<td>0.0900</td>
</tr>
<tr>
<td>$BPV_{5s}$</td>
<td>0.6408</td>
<td>0.3464</td>
<td>2.7453</td>
<td>19.273</td>
<td>0.5558</td>
<td>0.0949</td>
</tr>
<tr>
<td>$BPV_{5m}$</td>
<td>0.9691</td>
<td>0.7239</td>
<td>3.0619</td>
<td>18.300</td>
<td>0.5148</td>
<td>0.0381</td>
</tr>
</tbody>
</table>

Table 4: Sample statistics of $\tilde{r}_t = r_t / \sqrt{RV_{5m}}$, and realized measures of JPM.

5.1 TFSV model with leverage

We propose to use the observations on the RV series to carry out inference on the following TFSV model

$$dp^*(t) = \mu dt + \sigma_1(t) dW_1(t) + \sigma_2(t) dW_2(t)$$

$$d\sigma_1^2(t) = \kappa_1(\omega - \sigma_1^2(t)) dt + \varsigma_1 \sigma_1(t) dW_3(t)$$

$$d\sigma_2^2(t) = \kappa_2(\omega - \sigma_2^2(t)) dt + \varsigma_2 \sigma_2(t) dW_4(t)$$

$$\text{Corr}(dW_1(t), dW_3(t)) = \rho_1$$

$$\text{Corr}(dW_2(t), dW_4(t)) = \rho_2$$

where the parameters $\kappa_1$ and $\kappa_2$ govern the speed of mean reversion, while $\varsigma_1$ and $\varsigma_2$ determine the volatility of the volatility innovations. The parameter $\omega$ is the long-run mean of each volatility component and, as in Corsi and Renò (2012), it is assumed to be the same for both $\sigma_1^2(t)$ and $\sigma_2^2(t)$, in order to guarantee identification. $\{W_1(t) : t \geq 0\}, \{W_2(t) : t \geq 0\}, \{W_3(t) : t \geq 0\}, \{W_4(t) : t \geq 0\}$ are standard Brownian motions and $p^*(t)$ denotes the efficient log-price. The leverage effect is induced by the parameters $\rho_1$ and $\rho_2$, while the MN is modeled either as an i.i.d Gaussian variable that is added to the log-returns as in Bandi and Russell (2006), $r_{t,i} = r_{t,i}^* + \nu_{t,i}$, or as the bid-ask spread. In particular, the bid-ask bounce is generated as

$$p_{t,i} = p_{t,i}^* + \frac{\xi}{2} \Pi_{t,i}$$

(44)
where $\xi$ is the spread, and the order-driven indicator variables $I_{t,i}$ are independently across $t$ and $i$ and identically distributed with $Pr\{I_{t,i} = 1\} = Pr\{I_{t,i} = -1\} = \frac{1}{2}$. This variable takes value 1 when the transaction is buyer-initiated, and $-1$ when it is seller-initiated. Depending on the contamination scheme adopted, the structural parameters are collected in the vectors $\zeta_G = [\kappa_1, \kappa_2, \omega, \varsigma_1, \varsigma_2, \mu, \rho_1, \rho_2, \sigma^2_{\omega}]'$ and $\zeta_{BA} = [\kappa_1, \kappa_2, \omega, \varsigma_1, \varsigma_2, \mu, \rho_1, \rho_2, \xi]'$, where both $\zeta_G$ and $\zeta_{BA}$ are $(9 \times 1)$ vectors.

Since the leverage effect in the structural TFSV model is given by the contemporaneous dependence between the innovations of the returns and volatility processes, to identify it we need a multivariate auxiliary model that takes into account the contemporaneous covariance between the innovations to the volatility and log-return equations. Therefore, differently from Corsi and Renò (2012) that add past negative returns as explanatory variables in the HAR model, our auxiliary model is based on daily observations of $\tilde{r}_t$ and $RV$ for $t = 1, ..., T$, and it is given by the following tri-variate process

$$
\tilde{r}_t = \phi_0 + \phi_1 \tilde{r}_{t-1} + e^\prime_t,
$$

$$
\log(RV^5_{t,m}) = \beta_{0,m} + \beta_{1,m} \log(RV^5_{t-1,m}) + \beta_{2,m} \log(RV^5_{t-1,w}) + \beta_{3,m} \log(RV^5_{t-1,m}) + e^m_t,
$$

$$
\log(RV^5_{t,s}) = \beta_{0,s} + \beta_{1,s} \log(RV^5_{t-1,m}) + \beta_{2,s} \log(RV^5_{t-1,w}) + \beta_{3,s} \log(RV^5_{t-1,m}) + e^s_t,
$$

$$
\Sigma = \text{Cov}([e^e_t, e^m_t, e^s_t])
$$

The auxiliary parameters are $\beta = [\phi_0, \phi_1, \beta_{0,m}, \beta_{1,m}, \beta_{2,m}, \beta_{3,m}, \beta_{0,s}, \beta_{1,s}, \beta_{2,s}, \beta_{3,s}, \text{vech}(\Sigma)]'$ and $\beta$ is a $(16 \times 1)$ vector. As in Corsi and Renò (2012), the weighting matrix, $\Omega$, is chosen as the inverse of the asymptotic covariance matrix of the auxiliary estimates, $\Omega = \text{Var}(\hat{\beta})^{-1}$.

First, the TFSV model is estimated assuming that the drift and the leverage are both absent. The estimates of this restricted case are contained in Table 5. The first two lines report the estimates of the structural parameters based on univariate HAR models on log($RV^5_{t,m}$) and log($RV^5_{t,s}$) respectively with no correction for the noise. In this case, there are 5 auxiliary parameters and 5 structural parameters, so that the model is exactly identified. Notably, the $\chi^2$ criterion function is zero, meaning that the fitting to the data is perfect. However, the estimates of the parameters strongly depend on the sampling frequency selected to compute $RV$. Indeed, the long-run mean $\omega$, is approximately 30% higher for $RV^{5s}$ than $RV^{5m}$, reflecting the differences observed in the sample statistics. The speed of mean reversion $\kappa_2$ is two time lower when $RV^{5s}$ is used instead of $RV^{5m}$ reflecting the higher persistence of $RV^{5s}$. When the bivariate HAR model is estimated, but the presence of the MN is neglected, the criterion function is minimized very far from zero at $\chi^2 = 126.9$. This means that it is not possible to match the moments arising from the two $RV$ series unless the presence of the MN is explicitly accounted for.

When the noise is modeled as an i.i.d. Gaussian random variable with variance $\sigma^2_{\omega}$ or with bid-ask spread of size $\xi$, the adherence of the simulated processes to the observed $RV$ series improves significantly and the $\chi^2$ only leads to a marginal rejection of the model. Looking at the estimates of the parameters, we notice they are almost identical both under i.i.d. Gaussian distribution and bid-ask spread. In particular, the long-run mean $\omega$ is around 0.46 and the other TFSV parameters are generally within the estimates obtained when the HAR model on $RV^{5m}$ and $RV^{5s}$ are estimated separately. Moreover, the estimates of $\sigma^2_{\omega}$ and $\xi$ are almost identical, indicating that the two perturbation schemes have similar effects in the contamination of the latent efficient process.

The results do not change substantially when the model with leverage effect is estimated, as it emerges from Table 6. In this case, the auxiliary model contains the equation of the $\tilde{r}_t$, and the covariance matrix $\Sigma$ should provide enough information to identify the leverage parameters. The estimates of $\rho_1$ are approximately $-0.30$, while the estimate of $\rho_2$ is positive for the model with NO NOISE and BID-ASK, and negative for the model with Gaussian noise. However the estimates

\[^8\]In this case, the vectors of structural parameters are $\zeta_G = [\kappa_1, \kappa_2, \omega, \varsigma_1, \varsigma_2, \sigma^2_{\omega}]'$ and $\zeta_{BA} = [\kappa_1, \kappa_2, \omega, \varsigma_1, \varsigma_2, \xi]'$, and the auxiliary model does not include the equation with $\tilde{r}_t$, and it has 11 free parameters. In this restricted case, $p - q = 6$. 

22
\[
\begin{array}{cccccccc}
\kappa_1 & \kappa_2 & \omega & \varsigma_1 & \varsigma_2 & \sigma_{11}^2 & \xi & \chi^2 \\
\hline
\text{UNIV (5 min)} & 2.3306^a & 0.0559^a & 0.4905^a & 1.7690^a & 0.1934^a & - & - & 0.000^* \\
\text{UNIV (5 sec)} & 2.9712^a & 0.0279^b & 0.6850^a & 1.4301^a & 0.1351^a & - & - & 0.000^* \\
\text{NO NOISE:} & 2.4659^a & 0.0232^a & 0.6239^a & 1.3117^a & 0.1388^a & - & - & 126.9^a \\
\text{BID-ASK:} & 2.6781^a & 0.0396^a & 0.4698^a & 1.7755^a & 0.1674^a & - & 0.0001^a & 10.94^c \\
\text{GAUSS:} & 2.6917^a & 0.0394^a & 0.4699^a & 1.7849^a & 0.1668^a & 0.0001^a & - & 11.06^c \\
\end{array}
\]

Table 5: Estimates of the TF Heston parameters with indirect inference. \(a, b, c\) and \(d\) stand for significance at 1%, 5% and 10% respectively. Last column reports the value of the criterion function \(\chi^2\). The superscript * indicates that models UNIV (5 min) and UNIV (5 sec) are exactly identified, as \(p = q\).

of \(\rho_1, \rho_2\) and \(\mu\) turn out to be non significant. This is probably due to the fact that the sample period 2003-2007 is not affected by financial crises, thus making the leverage effect less relevant than in other periods. The fact that the criterion function \(\chi^2\) is not zero could signal some degree of misspecification in the structural model or in the added MN term, so that the parameters of the equations \(RV^{5m}\) and \(RV^{5s}\) do not match perfectly with those estimated on the data. Therefore, a closer look at the auxiliary estimates may be very informative. Indeed, the causes of the inability of the structural model to completely fit the data arise by looking at the parameter estimates of the auxiliary models. It emerges an overall good matching of the auxiliary parameters when the bid-ask and the Gaussian noise are adopted for the contamination. However, the estimated value of the parameter \(\sigma_{11}^2\), that is the variance of the innovation in the \(\tilde{r}_t\) equation, does not match with that estimated on the observed series. Indeed, the structural model implies that \(\sigma_{11}^2\) is very close to 1, as the only source of the movements in the returns is the volatility. On the contrary, \(\sigma_{11}^2\) is estimated equal to 0.81 on the observed data, meaning that 20% of the variation of returns cannot be attributed to the volatility. In other words, \(RV^{5m}\) produces an over-estimation of the return variation as generated by the continuous volatility trajectories. This evidence calls for an extension of the structural model, to account for other sources of the total return variation. A straightforward extension is to account for the possibility of abrupt movements in the log-price dynamics, as those generated by price jumps.

\[
\begin{array}{ccccccc}
\kappa_1 & \kappa_2 & \omega & \varsigma_1 & \varsigma_2 & \chi^2 \\
\hline
\text{NO NOISE:} & 2.5232a & 0.0230 & 0.6198a & 1.3241a & 0.1405^a \\
\text{BID-ASK:} & 2.7149a & 0.0380 & 0.4703a & 1.7784a & 0.1687^a \\
\text{GAUSS:} & 2.7555a & 0.0405 & 0.4719a & 1.8015a & 0.1696^a \\
\end{array}
\]

Table 6: Estimates of the TFSV model with leverage. \(a, b, c\) and \(d\) stand for significance at 1%, 5% and 10% respectively. Last column reports the value of the criterion function \(\chi^2\).

\footnote{For example, the MN term could be made more realistic, letting it follow an AR(1) process or letting its variance to follow a GARCH process. This poses the problem of developing a new auxiliary model, that is able to identify the extra nuisance parameters of the MN term. This is left to future research.}
with $\tilde{r}_t = r_t / \sqrt{RV_{t,5m}^5}$. This model employs the information arising from $BPV_{t,5s}^5$ to disentangle the effect of price jumps from the continuous volatility trajectories. Indeed, following the argument in Christensen et al. (2014), the squared price jumps can be well separated from the integrated variance only when $BPV_{t,5s}^5$ is computed using return at very high frequencies. Therefore, we expect that the combining $RV_{t,5m}^5$ and $BPV_{t,5s}^5$ will provide sufficient information to identify the MN parameters as well as the jump parameters. The set of auxiliary parameters is therefore $\beta = [\phi_0, \phi_1, \beta_{0,m}, \beta_{1,m}, \beta_{2,m}, \beta_{3,m}, \beta_{0,s}, \beta_{1,s}, \beta_{2,s}, \beta_{3,s}, \text{vech}(\Sigma)]'$. Table 8 reports the parameter estimates of model TFSV-J under different auxiliary models and assumptions on the noise. The univariate estimates, UNIV (5 min) and UNIV (5 sec), are based on an auxiliary model similar to that adopted in Corsi and Renò (2012)
where \( r_{t-1}, r_{t-1,w}, r_{t-1,m} \) are the past negative log-returns cumulated over daily, weekly and monthly horizons, and \( J_t = RV_t - BPV_t \) is a proxy for the squared jumps term. In model UNIV (5 min), \( RV_t \) and \( BPV_t \) are based on log-returns sampled every 5 minutes, while in model UNIV (5 sec) they are based on log-returns computed on grids of 5 seconds. The parameter estimates and the fitting are dramatically affected by the sampling frequency adopted in the estimation. Interestingly, the estimates of \( \lambda \), which measures the average number of jumps per day, are very high in both cases, but the jump sizes, as measured by \( \sigma^2_J \), are almost null, meaning that the price jump component is statistically insignificant. In particular, sampling too sparsely, i.e. at 5 minutes, does not provide sufficient information to disentangle price jumps from discretization error and volatility, thus resulting in insignificant estimates of the jumps, even if the overall fitting is quite good as \( \chi^2 = 6.899 \). When sampling more frequently, the volatility signal is strongly affected by the MN, thus resulting in biased estimates of \( \omega \), negligible estimates of the jump term and poor fitting, since \( \chi^2 \) is above 40. Moreover, the standard errors are very high for almost all parameters, so that only few of them are found significant. This may signal poor identification of all model parameters.

<table>
<thead>
<tr>
<th></th>
<th>( \kappa_1 )</th>
<th>( \kappa_2 )</th>
<th>( \omega )</th>
<th>( \varsigma_1 )</th>
<th>( \varsigma_2 )</th>
<th>( \mu )</th>
</tr>
</thead>
<tbody>
<tr>
<td>UNIV (5 min)</td>
<td>3.7082\textsuperscript{\textit{a}}</td>
<td>0.0811\textsuperscript{\textit{a}}</td>
<td>0.4280</td>
<td>2.3712</td>
<td>0.2161</td>
<td>0.0438</td>
</tr>
<tr>
<td>UNIV (5 sec)</td>
<td>5.3100</td>
<td>0.0042</td>
<td>1.5996</td>
<td>2.2868</td>
<td>0.4518</td>
<td>0.0035</td>
</tr>
<tr>
<td>NO NOISE</td>
<td>2.4552\textsuperscript{\textit{a}}</td>
<td>0.0226</td>
<td>0.4241\textsuperscript{\textit{a}}</td>
<td>1.1020\textsuperscript{\textit{a}}</td>
<td>0.1680\textsuperscript{\textit{a}}</td>
<td>0.2027\textsuperscript{\textit{a}}</td>
</tr>
<tr>
<td>BID-ASK</td>
<td>3.5882\textsuperscript{\textit{a}}</td>
<td>0.0373\textsuperscript{\textit{b}}</td>
<td>0.3501\textsuperscript{\textit{a}}</td>
<td>1.6470\textsuperscript{\textit{a}}</td>
<td>0.2220\textsuperscript{\textit{a}}</td>
<td>0.2927\textsuperscript{\textit{c}}</td>
</tr>
<tr>
<td>GAUSS</td>
<td>3.9083</td>
<td>0.0273\textsuperscript{\textit{a}}</td>
<td>0.3500\textsuperscript{\textit{a}}</td>
<td>1.7477\textsuperscript{\textit{c}}</td>
<td>0.2596\textsuperscript{\textit{c}}</td>
<td>0.2982\textsuperscript{\textit{c}}</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>( \rho_1 )</th>
<th>( \rho_2 )</th>
<th>( \lambda )</th>
<th>( \sigma^2_J )</th>
<th>( \sigma^2_\nu ) and ( \xi )</th>
<th>( \chi^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>UNIV (5 min)</td>
<td>0.1350</td>
<td>-0.5422</td>
<td>13.891\textsuperscript{\textit{a}}</td>
<td>0.0012</td>
<td>–</td>
<td>6.8993</td>
</tr>
<tr>
<td>UNIV (5 sec)</td>
<td>-0.1119</td>
<td>0.1542</td>
<td>40.248\textsuperscript{\textit{a}}</td>
<td>0.0015</td>
<td>–</td>
<td>40.488\textsuperscript{\textit{a}}</td>
</tr>
<tr>
<td>NO NOISE</td>
<td>-0.4213</td>
<td>0.5732</td>
<td>0.4680\textsuperscript{\textit{b}}</td>
<td>0.2020\textsuperscript{\textit{a}}</td>
<td>–</td>
<td>47.740\textsuperscript{\textit{a}}</td>
</tr>
<tr>
<td>BID-ASK</td>
<td>-0.3863</td>
<td>0.3139</td>
<td>3.7440\textsuperscript{\textit{a}}</td>
<td>0.0598\textsuperscript{\textit{a}}</td>
<td>0.0001\textsuperscript{\textit{a}}</td>
<td>46.223\textsuperscript{\textit{a}}</td>
</tr>
<tr>
<td>GAUSS</td>
<td>-0.3828\textsuperscript{\textit{a}}</td>
<td>0.3617</td>
<td>5.6160\textsuperscript{\textit{a}}</td>
<td>0.0497\textsuperscript{\textit{a}}</td>
<td>0.0001\textsuperscript{\textit{a}}</td>
<td>46.157\textsuperscript{\textit{c}}</td>
</tr>
</tbody>
</table>

Table 8: Estimates of the TFSV-J model with leverage. \( a \), \( b \) and \( c \) stand for significance at 1%, 5% and 10% respectively. Last column reports the value of the criterion function \( \chi^2 \).

When turning our attention to the estimates obtained with the multivariate model, the jump term is found significant in all cases since both \( \lambda \) and \( \sigma^2_J \), as well as their product, are significant, while the estimates of the other SV parameters are not much affected by the inclusion of jumps compared to the values reported in Table 6. The product \( \hat{\lambda} \hat{\sigma}^2_J \), which is a proxy of a jump average size, is significant and approximately 0.22 and 0.27 for BID-ASK and GAUSS cases respectively, meaning that part of the gap between the unconditional mean of \( BPV_{5s} \) and \( RV_{5m} \), see Table 4, is attributed to the jump term. Notably, the estimates of the long-run mean, \( \omega \), are smaller than those reported in Table 6 as the jump component is now responsible for a significant portion of the total return variation, such that the estimated unconditional variance is lower than when jumps are excluded from the model. Most parameters, with the exclusion of those governing the leverage, are significant. On the contrary, the fitting of the model is not optimal, as the criterion function is above 45, thus leading to a rejection of the structural model. An explanation for this poor performance is due to the inability of the contamination methods, either with i.i.d. Gaussian noise or bid/ask spread, to provide a realistic setup for the generation of high frequency returns. Indeed, the log-prices at very high frequencies are characterized by discreteness, due to the \textit{decimalization} and rounding effects, making the \( BPV_{5s} \) series severely downward biased, as it emerges from the sample statistics in Table 4. It is clear that both the i.i.d. Gaussian noise and the bid/ask

\textsuperscript{10}The product \( \hat{\lambda} \hat{\sigma}^2_J = 0.0167 \) has a standard error of 0.0202.
spread are incapable of accounting for this feature. This misspecification is responsible for the fact that the criterion function is minimized far from zero, and its value is only marginally lower than that obtained when noise is completely neglected (NO NOISE). This evidence signals the importance of the correct specification of the contamination term, to guarantee a good fitting. Unfortunately, generating log-prices under a decimalization scheme is not a viable solution in the indirect inference framework as it induces discreetness in the observed log-price thus masking the impact of changes in the SV parameters on the continuous dynamics and leading to problems of identification. As a consequence, the numerical Jacobian of the binding function, necessary for the calculation of the standard errors, contains many values near zero and it is almost singular. This is very informative and may indicate that alternative structural models, involving for example pure-jump Lévy processes as in Barndorff-Nielsen and Shephard (2001) or in Barndorff-Nielsen et al. (2014), may be better suited to model stock returns at very high-frequencies and could possibly be estimated by indirect inference.

6 Conclusions

The paper studies the inconsistency problem of indirect inference estimator caused by measurement error in the observed series. We show that this inconsistency is originated by a mismatch between the binding function implied by the observed data and that obtained by simulation. We propose a general method to deal with this error-in-variable problem in the indirect inference framework. The solution is to jointly estimate the nuisance parameters of the error term distribution and the structural ones. Hence, the simulated series used to match the auxiliary parameters must be contaminated by the noise. Under standard assumptions, this estimator is consistent and asymptotically normal. We provide a condition for the identification of ARMA model when contaminated by i.i.d. noise. This results may be helpful to check identification in economic and financial models that have an ARMA plus noise reduced form. One of these cases is represented by the Heston SV model. We prove that the HAR-RV auxiliary model guarantees the identification. Monte Carlo simulations show the viability of the proposed method in this framework and highlights the trade-off between bias reduction and efficiency as the sampling frequency changes. The empirical application illustrates the practical usefulness of the proposed methodology and the need for a correct specification of the conditional distribution of the microstructure noise term. A detailed analysis of the robustness/sensitivity of the indirect inference estimation to misspecifications of the measurement error is left to future research.

References


A Examples

A.1 Ornstein-Uhlenbeck

The exact discretized process of $z(t)$ is

$$y_t = \omega (1 - e^{-k}) + e^{-k} y_{t-1} + \sigma \left( \frac{1 - e^{-2k}}{2k} \right)^{1/2} \epsilon_t, \quad \epsilon_t \sim N(0,1),$$

with $E[y_t] = \omega$, $\text{Var}[y_t] = \frac{\sigma^2}{2k}$ and $E[y_t y_{t-1}] = \omega^2 + \frac{\sigma^2 e^{-k}}{2k}$. The auxiliary model can also be written as

$$x_t = \gamma_1 + \gamma_2 x_{t-1} + \gamma_3 \epsilon_t, \quad \epsilon_t \sim i.i.d.N(0,1)$$

with $\beta_1 = 1 - \gamma_2$, $\beta_2 = \gamma_1 / (1 - \gamma_2)$ and $\gamma_3 = \beta_3$. The maximum likelihood estimators of $\gamma_1$, $\gamma_2$ and $\gamma_3$ based on $x_t$, are

$$\hat{\gamma}_1 = \frac{1}{1 - \gamma_2} \sum_t x_t, \quad \hat{\gamma}_2 = \frac{1}{1 - \gamma_2} \sum_t x_t x_{t-1}, \quad \hat{\gamma}_3 = \frac{1}{1 - \gamma_2} \sum_t (x_t - \hat{\gamma}_1 - \hat{\gamma}_2 x_{t-1})^2.$$

The probability limit of $\hat{\beta}_T = (\hat{\beta}_1, T, \hat{\beta}_2, T, \hat{\beta}_3, T)'$ is readily obtained from $\hat{\gamma}$. Consider

$$p \lim \hat{\gamma}_2 = p \lim \frac{\sum_t x_t x_{t-1} - \sum_t x_t \sum_t x_{t-1}}{\sum_t x_t^2 - (\sum_t x_{t-1})^2}.$$

The limit in probability of the numerator is

$$p \lim \frac{1}{T} \sum_t x_t x_{t-1} = p \lim \frac{1}{T} \sum_t y_t y_{t-1} = \omega^2 + e^{-k} \frac{\sigma^2}{2k}$$

$$p \lim \frac{1}{T} \sum_t x_t \frac{1}{T} \sum_t x_{t-1} = \omega^2,$$

the limit of the denominator

$$p \lim \frac{1}{T} \left\{ \sum_t (y_{t-1} + u_{t-1})^2 - \left[ \sum_t (y_{t-1} + u_{t-1}) \right]^2 \right\} = \text{Var}[y_t + u_t] = \text{Var}[y_t] + \sigma_u^2.$$

Combining the two we obtain

$$p \lim \hat{\beta}_1 = 1 - p \lim \hat{\gamma}_2 = 1 - \frac{e^{-k} \sigma^2}{\sigma^2 + 2k \sigma_u^2},$$

Now, $\hat{\beta}_2$

$$\hat{\beta}_2 = \hat{\gamma}_1 \hat{\beta}_1$$

$$p \lim \hat{\gamma}_1 = \omega \hat{\beta}_1$$

thus

$$p \lim \hat{\beta}_2 = \hat{\gamma}_1 \hat{\beta}_1 = \omega.$$
Noting that $E[x_1] = \omega$, $E[x_1^2] = \text{Var}[y_t] + \omega^2 + \sigma_u^2$ and $E[x_1x_{t-1}] = e^{-k} \text{Var}(y_t) + \omega^2$. We also have to keep in mind that $\lim p \beta_1 = 1 - \frac{e^{-k}\sigma_u^2}{\sigma^2 + 2k\sigma^2_{u,0}}$. Hence, we get the plim of $\hat{\beta}_3$, that is

$$\lim p \hat{\beta}_3 = \left(\frac{\sigma^2}{2k} + \sigma_u^2\right)(1 + (1 - \beta_1)^2) - (1 - \beta_1)e^{-k}\left(\frac{\sigma^2}{k}\right).$$

Summarizing, let $\theta_0 = (k_0, \omega_0, \sigma_0)'$ and $\psi_0 = \sigma_{u,0}^2$, the probability limit of $\hat{\beta}$ is given by

$$b(\theta_0, \psi_0) = \left[\begin{array}{c} 1 - \frac{e^{-k}\sigma_u^2}{\sigma_0^2 + 2k_0\sigma^2_{u,0}} \\
\omega \\
\left(\frac{\sigma^2}{2k_0} + \sigma_{u,0}^2\right)[1 + \left(\frac{e^{-k_0}\sigma_u^2}{\sigma_0^2 + 2k_0\sigma^2_{u,0}}\right)^2] - \left(\frac{e^{-k_0}\sigma_u^2}{\sigma_0^2 + 2k_0\sigma^2_{u,0}}\right)e^{-k_0}\sigma_u^2 \right]^{1/2}
\end{array}\right] \quad (A.1)$$

While the function $\bar{b}(\theta)$ is

$$\bar{b}(\theta) = \left[\begin{array}{c} 1 - e^{-k} \\
\omega \\
\frac{1}{\sigma\omega^2}(1 - e^{-2k})^{1/2}
\end{array}\right]$$

with

$$\frac{\partial \bar{b}(\theta_0)}{\partial \theta} = \left[\begin{array}{ccc} e^{-k_0} & 0 & 0 \\
0 & 1 & 0 \\
\left(\frac{\sigma_{u,0}^2}{\sigma^2 + 2k_0\sigma^2_{u,0}}\right)^{1/2} & \left(\frac{1 - e^{-2k_0}}{2k_0}\right) & \left(\frac{1 - e^{-2k_0}}{2k_0}\right)^{1/2}
\end{array}\right].$$

### A.2 Identification of the MA(1) plus measurement error

The binding function is given by

$$b_\epsilon(\theta, \sigma_u^2) = \lim_{T \to \infty} \frac{\hat{\phi}_{2T}}{\hat{\phi}_{2T}}$$

$$\lim_{T \to \infty} \hat{\phi}_{2T} = \left[(1 + \varphi^2)\sigma_e^2 + \sigma_u^2 + \frac{\varphi^2 \sigma_e^2}{\varphi \sigma_u^2} (1 + \varphi^2)\sigma_e^2 + \sigma_u^2\right]^{-1} \left[\begin{array}{c} \varphi \sigma_u^2 \\
0
\end{array}\right]$$

The limit of the OLS estimator of $\phi_1, \phi_2$ are

$$\lim p \hat{\phi}_{1T} = \frac{1}{\left[(1 + \varphi^2)\sigma_e^2 + \sigma_u^2\right] + \varphi^2 \sigma_e^2}$$

$$\lim p \hat{\phi}_{2T} = \frac{-\varphi^2 \sigma_e^4}{\left[(1 + \varphi^2)\sigma_e^2 + \sigma_u^2\right] + \varphi^2 \sigma_e^2}.$$
B Proofs

Following (Gouriéroux and Monfort, 1996, p.78), the second order expansion of \( \hat{\zeta} \) is given by:

\[
p \lim \frac{1}{T} \sum_{t} \hat{\phi}_{1,T}^2 x_{t-1} = 2 \left[ \frac{[(1 + \varphi^2) \sigma_x^2 + \sigma_u^2] \varphi \sigma_x^2}{(1 + \varphi^2) \sigma_x^2 + \sigma_u^2} \right] \left[ \frac{-(1 + \varphi^2) \sigma_x^2 + \sigma_u^2}{(1 + \varphi^2) \sigma_x^2 + \sigma_u^2} \right] \varphi \sigma_x^2.
\]

The result is

\[
p \lim \frac{1}{T} \sum_{t} \hat{\phi}_{1,T}^2 x_{t-1} = (1 + \varphi^2) \sigma_x^2 + \sigma_u^2 + 3 \frac{[(1 + \varphi^2) \sigma_x^2 + \sigma_u^2]}{(1 + \varphi^2) \sigma_x^2 + \sigma_u^2 - \varphi^2 \sigma_x^4} \varphi \sigma_x^2.
\]

The Jacobian matrix \( \frac{\partial \hat{\theta}(\zeta)}{\partial \zeta} \) can be obtained in closed form evaluating the first derivative of the binding function for each element of the vector \( \zeta \).

B Proofs

B.1 Proof of Proposition 2.1

The first step estimator \( \hat{\beta}_T^*(\theta) \) associated with the set of simulated values \( \{v_s, s = 1, \ldots, S\} \)

\[
\hat{\beta}_T^*(\theta) = \hat{b}(\theta) + \frac{A(v_s; \theta)}{\sqrt{T}} + \frac{B(v_s; \theta)}{T} + o \left( \frac{1}{T} \right),
\]

(B.1)

where \( v_s \) is i.i.d. An indirect inference estimator \( \hat{\theta}_{ST} \) can be defined as the solution of the system

\[
\hat{\beta}_T(\theta_0, \psi_0) = \frac{1}{S} \sum_{s=1}^{S} \hat{\beta}_T^*(\hat{\theta}_{ST}).
\]

(B.2)

The second order expansion of \( \hat{\theta}_{ST} \)

\[
\hat{\theta}_{ST} = \theta_0 + \frac{a^*}{\sqrt{T}} + \frac{b^*}{T} + o \left( \frac{1}{T} \right)
\]

(B.3)

Following (Gouriéroux and Monfort, 1996, p.78)

\[
b(\theta_0, \psi_0) + \frac{A(v_s; \theta_0, \psi_0)}{\sqrt{T}} + \frac{B(v_s; \theta_0, \psi_0)}{T} =
\]

\[
\frac{1}{S} \sum_{s=1}^{S} \left\{ \hat{b}(\hat{b}_{ST}) + \frac{A(v_s; \hat{\theta}_{ST})}{\sqrt{T}} + \frac{B(v_s; \hat{\theta}_{ST})}{T} \right\} + o \left( \frac{1}{T} \right)
\]

(B.4)
In order to find the expression for $a^*$ and $b^*$ we have to expand in series $b(\hat{\theta}_{ST})$, $A(v_s; \hat{\theta}_{ST})$ and $B(v_s; \hat{\theta}_{ST})$ i.e.

\[
\tilde{b}(\hat{\theta}_{ST}) = \tilde{b}(\theta_0) + \frac{\partial \tilde{b}(\theta_0)}{\partial \theta'} (\hat{\theta}_{ST} - \theta_0) + \frac{1}{2} \left( I_q \otimes (\hat{\theta}_{ST} - \theta_0) \right)' F(\theta_0)(\hat{\theta}_{ST} - \theta_0) \tag{B.5a}
\]

\[
A(v_s; \hat{\theta}_{ST}) = A(v_s; \theta_0) + \frac{\partial A(v_s; \theta_0)}{\partial \theta'} (\hat{\theta}_{ST} - \theta_0) \tag{B.5b}
\]

\[
B(v_s; \hat{\theta}_{ST}) = B(v_s; \theta_0) + \frac{\partial B(v_s; \theta_0)}{\partial \theta'} (\hat{\theta}_{ST} - \theta_0), \tag{B.5c}
\]

where

\[
F(\theta) = [F_1(\theta)', \ldots, F_q(\theta)']' \]

with $F_i(\theta) = \frac{\partial b_i(\theta)}{\partial \theta'}$, where $\tilde{b}_i(\theta)$ is the i-th element of $\tilde{b}(\theta, \psi)$.

Inserting (B.5) in (B.4)

\[
b(\theta_0, \psi_0) + \frac{A(v_x; \theta_0, \psi_0)}{\sqrt{T}} + \frac{B(v_x; \theta_0, \psi_0)}{T} = \tilde{b}(\theta_0) + \frac{\partial \tilde{b}(\theta_0)}{\partial \theta'} \left( \frac{a^*}{\sqrt{T}} + \frac{b^*}{T} \right) + \frac{1}{2} \left( I_q \otimes (\hat{\theta}_{ST} - \theta_0) \right)' F(\theta_0)(\hat{\theta}_{ST} - \theta_0) \\
+ \frac{1}{S} \sum_{s=1}^{S} \left\{ \frac{1}{\sqrt{T}} \left[ A(v_s; \theta_0) + \frac{\partial A(v_s; \theta_0)}{\partial \theta'} \left( \frac{a^*}{\sqrt{T}} + \frac{b^*}{T} \right) \right] + \frac{B(v_s; \hat{\theta}_{ST}(\psi))}{T} \right\} \\
+ \frac{1}{ST} \sum_{s=1}^{S} \left\{ B(v_s; \theta_0) + \frac{\partial B(v_s; \theta_0)}{\partial \theta'} \left( \frac{a^*}{\sqrt{T}} + \frac{b^*}{T} \right) \right\} + o \left( \frac{1}{T} \right).
\]

Rearranging

\[
[b(\theta_0, \psi_0) - \tilde{b}(\theta_0)] + \frac{1}{\sqrt{T}} \left[ A(v_x; \theta_0, \psi_0) - \frac{1}{S} \sum_{s=1}^{S} A(v_s; \theta_0) \right] + \frac{1}{T} \left[ B(v_x; \theta_0, \psi_0) - \frac{1}{S} \sum_{s=1}^{S} B(v_s; \theta_0) \right] = \\
\frac{\partial \tilde{b}(\theta_0)}{\partial \theta'} \frac{a^*}{\sqrt{T}} + \frac{\partial \tilde{b}(\theta_0)}{\partial \theta'} \frac{b^*}{T} + \frac{1}{2} \left( I_q \otimes (\hat{\theta}_{ST} - \theta_0) \right)' F(\theta_0)(\hat{\theta}_{ST} - \theta_0) \\
+ \frac{1}{S} \sum_{s=1}^{S} \left\{ \frac{\partial A(v_s; \theta_0)}{\partial \theta'} \left( \frac{a^*}{\sqrt{T}} + \frac{b^*}{T} \right) \right\} + \frac{1}{ST} \sum_{s=1}^{S} \left\{ \frac{\partial B(v_s; \theta_0)}{\partial \theta'} \left( \frac{a^*}{\sqrt{T}} + \frac{b^*}{T} \right) \right\} + o \left( \frac{1}{T} \right)
\]

If the terms of order $o(T^{-1})$ are not explicitly considered

\[
[b(\theta_0, \psi_0) - \tilde{b}(\theta_0)] + \frac{1}{\sqrt{T}} \left[ A(v_x; \theta_0, \psi_0) - \frac{1}{S} \sum_{s=1}^{S} A(v_s; \theta_0) \right] + \frac{1}{T} \left[ B(v_x; \theta_0, \psi_0) - \frac{1}{S} \sum_{s=1}^{S} B(v_s; \theta_0) \right] \\
- \frac{1}{2} \left( I_q \otimes a^* \right)' F(\theta_0) a^* - \frac{1}{ST} \sum_{s=1}^{S} \frac{\partial A(v_s; \theta_0)}{\partial \theta'} a^* = \\
\frac{\partial \tilde{b}(\theta_0)}{\partial \theta'} \frac{a^*}{\sqrt{T}} + \frac{\partial \tilde{b}(\theta_0)}{\partial \theta'} \frac{b^*}{T} + o \left( \frac{1}{T} \right) \tag{B.6}
\]

Now, equating the members on both sides we get the expressions for $a^*$ and $b^*$ (see Gouriéroux and Monfort, 1996)

\[
a^* = \left[ \frac{\partial \tilde{b}(\theta_0)}{\partial \theta'} \right]^{-1} \left[ b(\theta_0, \psi_0) - \tilde{b}(\theta_0) \right] + \frac{1}{\sqrt{T}} \left[ \frac{\partial \tilde{b}(\theta_0)}{\partial \theta'} \right]^{-1} \left[ A(v_x; \theta_0, \psi_0) - \frac{1}{S} \sum_{s=1}^{S} A(v_s; \theta_0) \right] \tag{B.7}
\]
and

\[
b^\star = \left[ \frac{\partial \tilde{b}}{\partial \theta} (\theta_0) \right]^{-1} \left[ B(v_x; \theta_0) - \frac{1}{S} \sum_{s=1}^{S} B(v_s; \theta_0) \right] - \left[ \frac{\partial \tilde{b}}{\partial \theta} (\theta_0) \right]^{-1} \left\{ \frac{1}{S} \sum_{s=1}^{S} \frac{\partial A(v_s; \theta_0)}{\partial \theta} a^\star + \frac{1}{2} \left( I_q \otimes a^\star \right)' F(\theta_0) a^\star \right\} \tag{B.8}
\]

Since,

\[E[A(v_x; \theta_0)] \neq E[A(v_s; \theta_0)]\]

this implies that

\[E\left[ \frac{a^\star}{\sqrt{T}} \right] = \left[ \frac{\partial \tilde{b}(\theta_0)}{\partial \theta'} \right]^{-1} \left[ b(\theta_0, \psi_0) - \tilde{b}(\theta_0) \right] + \frac{1}{\sqrt{T}} \left[ \frac{\partial \tilde{b}(\theta_0)}{\partial \theta'} \right]^{-1} E\left[ A(v_x; \theta_0) - \frac{1}{S} \sum_{s=1}^{S} A(v_s; \theta_0) \right]. \tag{B.9}\]

While for \( b^\star \):

\[E[b^\star] = \left[ \frac{\partial \tilde{b}(\theta_0)}{\partial \theta} \right]^{-1} E\left[ B(v_x; \theta_0) - \frac{1}{S} \sum_{s=1}^{S} B(v_s; \theta_0) \right] - \left[ \frac{\partial \tilde{b}(\theta_0)}{\partial \theta} \right]^{-1} E\left\{ \frac{1}{S} \sum_{s=1}^{S} \frac{\partial A(v_s; \theta_0)}{\partial \theta} a^\star + \frac{1}{2} \left( I_q \otimes a^\star \right)' F(\theta_0) a^\star \right\} \tag{B.10}\]

B.2 Proof of Proposition 3.2

Let \( q = m + 2 \) and \( p + 1 = 3 + r + l \), with \( q \geq p + 1 \). Given the AR(m) model in (29), the OLS estimates of \( \beta = (\phi_0, \phi_1, \ldots, \phi_m) \) converges in probability, when \( T \) diverges, to \( Q_{ZZ}^{-1} Q_{ZX} \) (see Proposition 8.10.1 and Theorem 8.1.1 in Brockwell and Davis, 1991), while

\[p \lim_{T \to \infty} \hat{\sigma}_e^2 = Q_{XX} - Q_{XZ} Q_{ZZ}^{-1} Q_{ZX}\]

where

\[
Q_{ZZ} = \begin{bmatrix}
1 & \mu_x & \mu_x & \cdots & \mu_x \\
\mu_x & \gamma_x(0) + \mu_x^2 & \gamma_x(1) + \mu_x^2 & \cdots & \gamma_x(m - 1) + \mu_x^2 \\
\mu_x & \gamma_x(1) + \mu_x^2 & \gamma_x(0) + \mu_x^2 & \cdots & \gamma_x(m - 2) + \mu_x^2 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\mu_x & \gamma_x(m - 1) + \mu_x^2 & \gamma_x(m - 2) + \mu_x^2 & \cdots & \gamma_x(0) + \mu_x^2
\end{bmatrix}
\]

and \( Q_{XZ} = [\mu_x, \gamma_x(1) + \mu_x^2, \ldots, \gamma_x(m) + \mu_x^2] = Q_{ZX}' \). Since we assume that \( E(u_t) = 0 \), then \( \mu_x = \mu \equiv E[y_t] \). The variance and the autocovariances of \( x_t \) are \( Q_{XX} = \gamma_x(0) = \gamma(0) + \sigma_\theta^2 \) and \( \gamma_x(k) = \gamma(k) \) when \( k \neq 0 \), where \( \gamma(k) = \text{Cov}[y_t, y_{t-k}] \). Thus the binding function is

\[b(\zeta) = \begin{bmatrix} Q_{ZZ}^{-1} Q_{ZX} \\ Q_{XX} - Q_{XZ} Q_{ZZ}^{-1} Q_{ZX} \end{bmatrix}. \tag{B.11}\]

In order to find the Jacobian matrix of \( b(\zeta) \) consider the differential for each component of \( b(\zeta) \). Since

\[d(Q_{ZZ}^{-1} Q_{ZX}) = (dQ_{ZZ}^{-1}) Q_{ZX} + Q_{ZZ}^{-1} (dQ_{ZX})\]

\[= -(Q_{ZZ}^{-1} dQ_{ZZ} Q_{ZZ}^{-1}) Q_{ZX} + Q_{ZZ}^{-1} dQ_{ZX}\]

33
and given that $\text{vec}(ABCD) = (D'C' \otimes A)\text{vec}(B)$ for suitably dimensioned matrices, where the $\text{vec}$ operator transforms a matrix into a vector by stacking the columns of the matrix one underneath the other,

$$d\text{vec}(Q_{ZZ}^{-1}Q_{XX}) = -(Q_{XX}^{-1}Q_{ZZ})\text{vec}(Q_{ZZ}) + Q_{ZZ}^{-1}\frac{\partial Q_{XX}}{\partial \zeta'}d\zeta$$

$$= \left[ -(Q_{XX}^{-1}Q_{ZZ})\frac{\partial \text{vec}(Q_{ZZ})}{\partial \zeta'} + Q_{ZZ}^{-1}\frac{\partial Q_{XX}}{\partial \zeta'} \right]d\zeta$$

The differential of the second component of $b(\zeta)$ is

$$d[Q_{XX}] - d[Q_{XX}Q_{ZZ}^{-1}Q_{XX}] = d[Q_{XX}] - d[Q_{XX}Q_{ZZ}^{-1}Q_{XX}] = -\text{tr} \left( Q_{XX}(Q_{ZZ}^{-1}Q_{XX}) \right)$$

$$= -\text{vec}(Q_{XX}Q_{XX})'(Q_{ZZ}^{-1}Q_{ZZ})\frac{\partial \text{vec}(Q_{ZZ})}{\partial \zeta'}d\zeta$$

and

$$\{Q_{XX}Q_{ZZ}^{-1}d[Q_{XX}]\} = Q_{XX}Q_{ZZ}^{-1}\frac{\partial (Q_{XX})}{\partial \zeta'}d\zeta$$

Finally,

$$d[Q_{XX}] - d[Q_{XX}Q_{ZZ}^{-1}Q_{XX}] = \left[ \frac{\partial Q_{XX}}{\partial \zeta'} + \text{vec}(Q_{XX}Q_{XX})'(Q_{ZZ}^{-1}Q_{ZZ})\frac{\partial \text{vec}(Q_{ZZ})}{\partial \zeta'} - 2Q_{XX}Q_{ZZ}^{-1}\frac{\partial Q_{XX}}{\partial \zeta'} \right]d\zeta.$$
where
\[
\begin{align*}
\frac{\partial (\mu_x^2 + \gamma_x(k))}{\partial c} &= \frac{2c}{(1 - \sum_{i=1}^{r} \alpha_i)^2} \\
\frac{\partial (\mu_x^2 + \gamma_x(k))}{\partial \alpha_1} &= 2\mu \frac{\partial \mu}{\partial \alpha_1} + \frac{\partial \gamma(k)}{\partial \alpha_1} \\
\frac{\partial (\mu_x^2 + \gamma_x(k))}{\partial \varphi'} &= \frac{\partial \gamma(k)}{\partial \varphi'} \\
\frac{\partial (\mu_x^2 + \gamma_x(k))}{\partial \sigma^2} &= \frac{\partial \gamma(k)}{\partial \sigma^2} \\
\frac{\partial (\mu_x^2 + \gamma_x(k))}{\partial \sigma^2_u} &= \frac{\partial \gamma_x(k)}{\partial \sigma^2_u} = \left\{ \begin{array}{ll} 1 & k = 0 \\ 0 & k \neq 0 \end{array} \right.
\end{align*}
\]

First, consider the case of ARMA(1,0). The rows of $C$ have the following expression
\[
c_k = \left[ \frac{\partial (\mu_x^2 + \gamma_x(0))}{\partial c} \quad \frac{\partial (\mu_x^2 + \gamma_x(0))}{\partial \alpha_1} \quad \frac{\partial (\mu_x^2 + \gamma_x(0))}{\partial \sigma^2} \quad \frac{\partial (\mu_x^2 + \gamma_x(0))}{\partial \sigma^2_u} \right]'.
\]
If $C$ has reduced rank then $Cw = 0$ for $w \neq 0$, i.e. there exists a vector $w = [w_1, w_2, w_3, w_4]'$ such that $c_k'w = 0$ for all rows of $C$. The partial derivatives in $c_k$ are
\[
\begin{align*}
\frac{\partial (\mu_x^2 + \gamma_x(0))}{\partial c} &= \frac{2c}{(1 - \alpha_1)^2} \\
\frac{\partial (\mu_x^2 + \gamma_x(0))}{\partial \alpha_1} &= 2\mu \frac{\partial \mu}{\partial \alpha_1} + (k\alpha_1^{k-1}\gamma(0) + \alpha_1^k \frac{\partial \gamma(0)}{\partial \alpha_1}) \\
\frac{\partial (\mu_x^2 + \gamma_x(0))}{\partial \sigma^2} &= \alpha_1 \frac{\partial \gamma(0)}{\partial \sigma^2} \\
\end{align*}
\]
The reduced rank condition implies
\[
w_1 \left( \frac{\partial (\mu_x^2 + \gamma_x(0))}{\partial c} \right) + w_2 \left( \frac{\partial (\mu_x^2 + \gamma_x(0))}{\partial \alpha_1} \right) + w_3 \left( \frac{\partial (\mu_x^2 + \gamma_x(0))}{\partial \sigma^2} \right) + w_4 = 0 \quad \text{for } k = 0
\]
and for $k > 0$
\[
w_1 \left( \frac{\partial (\mu_x^2 + \gamma_x(k))}{\partial c} \right) + w_2 \left( \frac{\partial (\mu_x^2 + \gamma_x(k))}{\partial \alpha_1} \right) + w_3 \left( \frac{\partial (\mu_x^2 + \gamma_x(k))}{\partial \sigma^2} \right) = 0.
\]
Equating the two expressions above
\[
w_2 \left[ k\alpha_1^{k-1}\gamma(0) + \alpha_1^k \left( \frac{\partial \gamma(0)}{\partial \alpha_1} - \frac{\partial \gamma_x(0)}{\partial \alpha_1} \right) \right] + w_3 \frac{\partial \gamma_x(0)}{\partial \sigma^2}(\alpha^k - 1) = w_4
\]
it is easy to see that $w_4$ varies with $k$ which implies that cannot exist any vector $w \neq 0$ which lies in the null column subspace of $C$. We conclude that in the case of ARMA(1,0) the column rank of $C$ is full. Now, we show that when $r = l = 1$, on the contrary, the rank of $C$ is smaller than $p + 1$. In this case it is easy to show that there exists a vector $w \neq 0$ such that $Cw = 0$. For ARMA(1,1) the rows of $\partial b(\zeta)/\partial \zeta'$ consist of
\[
\left[ \frac{\partial (\mu_x^2 + \gamma_x(k))}{\partial c} \quad \frac{\partial (\mu_x^2 + \gamma_x(k))}{\partial \alpha_1} \quad \frac{\partial (\mu_x^2 + \gamma_x(k))}{\partial \varphi'} \quad \frac{\partial (\mu_x^2 + \gamma_x(k))}{\partial \sigma^2} \quad \frac{\partial (\mu_x^2 + \gamma_x(k))}{\partial \sigma^2_u} \right],
\]
where the variance and the autocovariances are
\[
\gamma_x(0) = \sigma^2_2 \frac{1 + \varphi^2_1 + 2\alpha_1 \varphi_1}{1 - \alpha_1^2} + \sigma^2_u
\]

35
\[\gamma(k) = \sigma^2 \alpha^k \frac{(\alpha_1 + \varphi_1)(1 + \alpha_1 \varphi_1)}{1 - \alpha_1^2} \quad k > 1\]

with

\[
\frac{\partial \gamma(k)}{\partial \varphi_1} = \sigma^2 \alpha^k \frac{1 + \alpha_1 \varphi_1 + \alpha_1 (1 + \varphi_1)}{1 - \alpha_1^2}
\]

\[
\frac{\partial \gamma(k)}{\partial \sigma^2} = \alpha^k \frac{(1 + \alpha_1 \varphi_1)(1 + \alpha_1 \varphi_1)}{1 - \alpha_1^2}.
\]

The vector \(w\) that is orthogonal to the rows of \(C\) is

\[w = (0, 0, 1, -\sigma^2 (1 + \alpha_1 \varphi_1 + \alpha_1 (1 + \varphi_1)) / (\alpha_1 + \varphi_1)(1 + \alpha_1 \varphi_1), w_5)'
\]

where

\[w_5 = \sigma^2 (1 + \alpha_1 \varphi_1 + \alpha_1 (1 + \varphi_1)) \frac{\partial \gamma(0)}{\partial \sigma^2} - \frac{\partial \gamma(0)}{\partial \varphi_1}.
\]

For the case \(r < l\) an analogous argument shows that the column rank of \(C\) is reduced. Thus when \(r \leq l\) the rank of \(C\) is smaller than \(p + 1\).

### B.3 Proof of Proposition 4.1

The HAR-RV model can be written as an AR(22) with linear restrictions on the autoregressive parameters

\[x_t = \alpha_0 + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \ldots + \alpha_{22} x_{t-22} + \epsilon_t \quad (B.12)\]

where \(\alpha_0 = \phi_1, \alpha_1 = (\phi_2 + \phi_3/5 + \phi_4/22), \alpha_2, ..., \alpha_5 = (\phi_3/5 + \phi_4/22)\) and \(\{\alpha_6, ..., \alpha_{22}\} = (\phi_4/22)\).

Let \(R\) be the \(23 \times 4\) matrix with the linear restrictions, then a compact expression for \(\alpha\) is \(\alpha = R\hat{\phi}\).

The restricted AR(22) model in (B.12) can be estimated by OLS imposing the restriction contained in the matrix \(R\). The \((23 \times 4)\) matrix \(R\) is

\[
R = \begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & \frac{1}{2} & \frac{1}{22} \\
0 & 0 & \frac{1}{2} & \frac{1}{22} \\
0 & 0 & \frac{1}{2} & \frac{1}{22} \\
0 & 0 & \frac{1}{2} & \frac{1}{22} \\
0 & 0 & 0 & \frac{1}{22} \\
0 & 0 & 0 & \frac{1}{22} \\
0 & 0 & 0 & \frac{1}{22} \\
\vdots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & \frac{1}{22} \\
\end{bmatrix} \quad (B.13)
\]

The OLS estimate of \(\phi\) is

\[
\hat{\phi}_T = \left[R' \left(\sum_t z_t z_t'\right) R\right]^{-1} R' \sum_t (z_t x_t)
\]

where \(z_t = (1, x_{t-1}, \ldots, x_{t-22})'\). Under standard assumptions, it can be shown that \(\frac{1}{T} \sum_t (z_t z_t') \overset{P}{\to} E[z_t z_t'] \equiv Q_{ZZ}\), where

\[
Q_{ZZ} = \begin{bmatrix}
1 & \mu & \mu & \ldots & \mu \\
\mu & \gamma(0) + \mu^2 & \gamma(1) + \mu^2 & \ldots & \gamma(21) + \mu^2 \\
\mu & \gamma(1) + \mu^2 & \gamma(0) + \mu^2 & \ldots & \gamma(20) + \mu^2 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\mu & \gamma(21) + \mu^2 & \gamma(20) + \mu^2 & \ldots & \gamma(0) + \mu^2
\end{bmatrix}
\]

36
and \( \mu = E[x_t] \). The probability limit of \( \hat{\alpha} \) is therefore

\[
p \lim_{T \to \infty} \hat{\alpha}_T = Q_{XX}^{-1} Q_{ZX}
\]

\( Q_{ZX} = E[z_t x_t] = [\mu, \mu^2 + \gamma(1), \ldots, \mu^2 + \gamma(22)]' \) and \( Q_{XX} = E[x_t^2] = \gamma(0) + \mu^2 \). The limit of \( \hat{\phi}_T \)

\[
p \lim_{T \to \infty} \hat{\phi}_T = \left[ R'p \lim_{T \to \infty} \left( \sum_t z_t'(z_t') R \right) \right]^{-1} R'p \lim_{T \to \infty} \sum_t (z_t x_t)
\]

\[
= [R'Q_{ZZ} R]^{-1} R'Q_{ZX}
\]

The matrix \( Q_{ZZ} \) and the vector \( Q_{ZX} \) both depend on the structural parameters \( \zeta \). The estimator of the variance of \( e_t \) is

\[
\hat{\sigma}^2_{eT} = \frac{\sum_t e_t^2}{T}
\]

\[
p \lim_{T \to \infty} \hat{\sigma}^2_{eT} = Q_{XX} - Q_{XZ} R (R'Q_{ZZ} R)^{-1} R'Q_{ZX}.
\]

Then the binding function results to be

\[
b(\zeta) = \left[ \frac{[R'Q_{ZZ} R]^{-1} R'Q_{ZX}}{Q_{XX} - [Q_{XZ} R (R'Q_{ZZ} R)^{-1} R'Q_{ZX}]} \right] \]

To calculate the derivative of \( b(\zeta) \) with respect to \( \zeta \) we use the differential of \( b(\zeta) \):

\[
d\left( [R'Q_{ZZ} R]^{-1} R'Q_{ZX} \right) = d\left( [R'Q_{ZZ} R]^{-1} \right) R'Q_{ZX} + [R'Q_{ZZ} R]^{-1} R'd\{Q_{ZX} \} \quad (B.14)
\]

The differential of first term in the RHS of \( (B.14) \)

\[
d\left( [R'Q_{ZZ} R]^{-1} \right) R'Q_{ZX} = -(R'Q_{ZZ} R)^{-1} d\{Q_{ZZ} \} R (R'Q_{ZZ} R)^{-1} R'Q_{ZX}
\]

\[
= -(R'Q_{ZZ} R)^{-1} R'd\{Q_{ZZ} \} R (R'Q_{ZZ} R)^{-1} R'Q_{ZX},
\]

taking the vec of both sides

\[
\text{vec}\left[d\left( [R'Q_{ZZ} R]^{-1} \right) R'Q_{ZX} \right] = \text{vec}\left[-(R'Q_{ZZ} R)^{-1} R'd\{Q_{ZZ} \} R (R'Q_{ZZ} R)^{-1} R'Q_{ZX} \right]
\]

\[
= -\left[ Q_{XZ} R (R'Q_{ZZ} R)^{-1} R' \otimes (R'Q_{ZZ} R)^{-1} R' \right] d\text{vec}(Q_{ZZ})
\]

\[
= -\left[ Q_{XZ} R (R'Q_{ZZ} R)^{-1} R' \otimes (R'Q_{ZZ} R)^{-1} R' \right] \frac{\partial \text{vec}Q_{ZZ}}{\partial \zeta'} d\text{vec} \zeta.
\]

Thus

\[
d\{R'Q_{ZZ} R\}^{-1} R'Q_{ZX} = \left\{ -\left[ Q_{XZ} R (R'Q_{ZZ} R)^{-1} R' \otimes (R'Q_{ZZ} R)^{-1} R' \right] \frac{\partial \text{vec}Q_{ZZ}}{\partial \zeta'} + (R'Q_{ZZ} R)^{-1} R' \frac{\partial Q_{ZX}}{\partial \zeta'} \right\} d\zeta.
\]

Now, the differential of the last row in \( b(\zeta) \)

\[
d\{Q_{XX} - [Q_{XZ} R (R'Q_{ZZ} R)^{-1} R'Q_{ZX}] \} = d\{Q_{XX} \} - d\{[Q_{XZ} R (R'Q_{ZZ} R)^{-1} R'Q_{ZX}] \}
\]

with

\[
d\{Q_{XX} \} = \frac{\partial Q_{XX}}{\partial \zeta'} d\zeta
\]
The Jacobian matrix of b matrix (as in the Proof of Proposition 3.2) and for \( j > C \) and this must hold for all rows of \( \gamma(j) \), i.e. 
\[
c_j = \left[ \frac{\partial \gamma(j)}{\partial \kappa} \ 2\mu + \frac{\partial \gamma(j)}{\partial \omega} \ \frac{\partial \gamma(j)}{\partial \kappa} \ 2\mu \Delta + \frac{\partial \gamma(j)}{\partial \sigma^2} \right]'.
\]

The matrix \( C \) has reduced column rank if there exists a vector \( w = [w_1, w_2, w_3, w_4]' \neq 0 \) such that
\[
e_j'w = 0
\]
and this must hold for all rows of \( C' \), that is \( Cw = 0 \). Since
\[
\frac{\partial \gamma(j)}{\partial \sigma^2} = 0 \quad \text{for } j > 0
\]
we have that for \( j > 0 \)
\[
w_1 \frac{\partial \gamma(j)}{\partial \kappa} + w_2 \left( 2\mu + \frac{\partial \gamma(j)}{\partial \omega} \right) + w_3 \left( 2\mu \Delta + \frac{\partial \gamma(j)}{\partial \sigma^2} \right) = 0 \quad \text{(B.15)}
\]
and for \( j = 0 \)
\[
w_1 \frac{\partial \gamma(0)}{\partial \kappa} + w_2 \left( 2\mu + \frac{\partial \gamma(0)}{\partial \omega} \right) + w_3 \frac{\partial \gamma(0)}{\partial \sigma^2} + w_4 \left( 2\mu \Delta + \frac{\partial \gamma(0)}{\partial \sigma^2} \right) = 0. \quad \text{(B.16)}
\]
Now equating (B.15) and (B.16), we get

\[
w_1 \left[ \frac{\partial \gamma(j)}{\partial \kappa} - \frac{\partial \gamma(0)}{\partial \kappa} \right] + w_2 \left[ \frac{\partial \gamma(j)}{\partial \omega} - \frac{\partial \gamma(0)}{\partial \omega} \right] + w_3 \left[ \frac{\partial \gamma(j)}{\partial \zeta} - \frac{\partial \gamma(0)}{\partial \zeta} \right] - w_4 \frac{\partial \gamma(0)}{\partial \sigma_\nu^2} = 0. \quad \text{(B.17)}
\]

where the expressions in parenthesis are functions of \( j \). This means that the value of \( w_4 \) which satisfies (B.17) is

\[
w_4 = \left[ \frac{\partial \gamma(0)}{\partial \sigma_\nu^2} \right]^{-1} \left\{ w_1 \left[ \frac{\partial \gamma(j)}{\partial \kappa} - \frac{\partial \gamma(0)}{\partial \kappa} \right] + w_2 \left[ \frac{\partial \gamma(j)}{\partial \omega} - \frac{\partial \gamma(0)}{\partial \omega} \right] + w_3 \left[ \frac{\partial \gamma(j)}{\partial \zeta} - \frac{\partial \gamma(0)}{\partial \zeta} \right] \right\}
\]

which is obviously not constant since it depends on \( j \). Thus the only vector \( w \) which satisfies (B.17) is the null vector. We can conclude that the matrix \( C \) has full column rank.