A Continuous-Time, Latent-Variable Model of Time-Series Data

Supplementary appendix

Contents

A. LIKELIHOOD COMPUTATION

A.1. *Kalman filter background*

I begin by giving a very brief overview of the basic Kalman filter algorithm. For a more thorough treatment of the Kalman filter, see Beck (1989). Readers familiar with the Kalman filter can jump to A.2, which describes likelihood computation of MONOCAR models based on this algorithm.

The Kalman filter is a recursive algorithm for estimating latent states from noisy observations in a linear discrete-time system. The Kalman filter was presented by Kalman (1960) and applied to likelihood computation by Schweppe (1965). It has received numerous applications across many fields and been described as "among the most notable innovations of the 20th century" (Humpherys, Redd, and West 2012).[1]

The Kalman filter assumes a model with a latent process, giving the evolution of an unobserved state vector, and an observation process, relating this latent process to observations. The latent process described by

$$\begin{aligned} \mathbf{y}_i &= \mathbf{\Phi}_i \mathbf{y}_{i-1} + \mathbf{B}_i \mathbf{u}_i + \boldsymbol{\varepsilon}_i \\ \varepsilon_i &\sim \mathcal{N}(0, \mathbf{T}_i) \end{aligned} \tag{A1}$$

where $\mathbf{y}_i$ is the unobserved value of the latent process at time i, $\mathbf{u}_i$ is a vector of exogenous variables (input variables in engineering contexts), ε_i is the error term for the latent process at time i, $\mathbf{T}_i$ is the variance-covariance matrix of the error term at time i, $\mathbf{\Phi}_i$ is a transition matrix defining how $\mathbf{y}_{i-1}$ affects the expectation of $\mathbf{y}_i$, and $\mathbf{B}_i$ is a similar matrix defining how $\mathbf{u}_i$ affects the expectation of $\mathbf{y}_i$. $\mathbf{B}_i$ and $\mathbf{u}_i$ are often omitted. Note that if $\mathbf{\Phi}_i$ and $\mathbf{T}_i$ do not vary with i and $\mathbf{B}_i$ and $\mathbf{u}_i$ are omitted, the resulting equation describes a first-order

[1] For a thorough treatment of the history of the algorithm, see Humpherys, Redd, and West (2012).

Gaussian VAR process.

The observation process is described by

$$\mathbf{z}_i = \mathbf{H}_i \mathbf{y}_i + \boldsymbol{\eta}_i \tag{A2}$$
$$\boldsymbol{\eta}_i \sim \mathcal{N}(0, \boldsymbol{\Psi}_i)$$

where $\mathbf{z}_i$ is the observation at time i, $\boldsymbol{\eta}_i$ is the error term for the observation process at time i, $\boldsymbol{\Psi}_i$ is the variance-covariance matrix of the error term at time i, and $\mathbf{H}_i$ defines how the latent process relates to the expected value of the observation.

The Kalman filter algorithm is a two-stage iterative process. Each iteration of the algorithm consists of a prediction stage, in which estimates of the state space at one time step are formed from estimates at the previous time step, and an update stage, in which the estimates of the state space are updated to reflect the observation for that time step. Each observation's contribution to the log-likelihood conditional on all previous observations can also be computed along with the updated stage. These contributions can then be summed to give the log-likelihood for the model. This algorithm is summarized by algorithm A1.

A range of models can be represented in the form described by equations A1 and A2, known as "state space form," with appropriate choices of $\boldsymbol{\Phi}_i$, $\mathbf{B}_i$, $\mathbf{T}_i$, $\mathbf{H}_i$, and $\boldsymbol{\Psi}_i$. ARMA models with measurement error and dynamic factor models are among the most common (see Beck 1989). As a result, the Kalman filter can be applied to estimating the log-likelihood for these models.

A.2. *Algorithm for MONOCAR likelihood computation*

In this section, I describe an algorithm for computing the log-likelihood of a MONOCAR model based on the Kalman filter. The algorithm is presented by algorithm A2, with some steps using results from section B. While implementing the MONOCAR likelihood

Algorithm A1: standard Kalman filter with log-likelihood computation

Input:

T = the number of time steps,
$\mathbf{x}_{0|0}$ = initial mean,
$\mathbf{V}_{0|0}$ = initial variance-covariance matrix, and
for each $i = 1$ **to** T
  $\mathbf{z_i}$, $\mathbf{\Phi}_i$, $\mathbf{B}_i$, $\mathbf{T}_i$, $\mathbf{H}_i$, and $\mathbf{\Psi}_i$
end

Result:

ℓ, the log-likelihood of the model, and
for each $i = 1$ **to** T
  $\mathbf{x}_{i|i}$ and $\mathbf{V}_{i|i}$
end

begin

$\ell \leftarrow 0$
for $i = 1$ **to** T **do**
  Note: `Prediction stage`
  $\mathbf{x}_{i|i-1} \leftarrow \mathbf{\Phi}_i \mathbf{x}_{i-1|i-1} + \mathbf{B}_i \mathbf{u}_i$
  $\mathbf{V}_{i|i-1} \leftarrow \mathbf{\Phi}_i \mathbf{V}_{i-1|i-1} \mathbf{\Phi}_i^\top + \mathbf{T}_i$
  Note: `Update stage`
  $\boldsymbol{\omega}_i \leftarrow \mathbf{z}_i - \mathbf{H}_i \mathbf{x}_{i|i-1}$
  $\Phi_i \leftarrow \mathbf{H}_i \mathbf{V}_{i|i-1} \mathbf{H}_i^\top + \mathbf{\Psi}_i$
  $\mathbf{W}_i \leftarrow \mathbf{V}_{i|i-1} \mathbf{H}_i^\top \Phi_i^{-1}$
  $\mathbf{x}_{i|i} \leftarrow \mathbf{x}_{i|i-1} + \mathbf{W}_i \boldsymbol{\omega}_i$
  $\mathbf{V}_{i|i} \leftarrow \mathbf{V}_{i|i-1} - \mathbf{W}_i \mathbf{H}_i \mathbf{V}_{i|i-1}$
  Note: `Likelihood update`
  $\ell \leftarrow \ell - \boldsymbol{\omega}_i^\top \Phi_i^{-1} \boldsymbol{\omega}_i - \frac{1}{2} \log |\Phi| - \frac{d}{2} \log (2\pi)$
end

end

computation is non-trivial, the R package developed for MONOCAR provides a optimized implementation of the algorithm with some extensions that can be applied by researchers.

Because MONOCAR can also be put in state space form, the Kalman filter can be applied to computing the log-likelihood of a MONOCAR model. However, computing the appropriate choices of the parameters used by the Kalman filter algorithm, such as $\mathbf{\Phi}_i$, for each step i is more complex than for many models. The dimensionality of the state space used by a MONOCAR model in state space form also varies between steps.[2] Some simplifications of the Kalman filter algorithm are also possible to streamline likelihood computation.

The state space used by the Kalman filter cannot simply be the state space of the model—that is, the space of possible values of $\mathbf{y}_t$—because this would be insufficient for non-instantaneous observations, as an observation recorded from time t_1 to time t_2 depends on $\mathbf{y}_{t'}$ for all $t_1 \leq t' \leq t_2$ rather than on $\mathbf{y}_{t'}$ for a single value of t'.

Although a non-instantaneous observation depends on the latent process at infinitely many points in time, it is possible to capture this dependency using a finite number of quantities. We can do so by adding components of the form

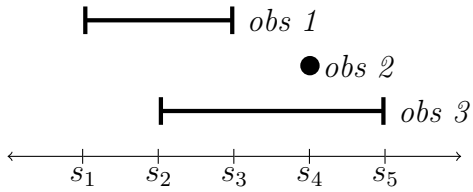
$$\bar{\mathbf{y}}_{t_1,t_2} = \int_{t_1}^{t_2} \mathbf{y}_s ds. \tag{A3}$$

Because observations may overlap, however, multiple $\bar{\mathbf{y}}$ may be needed at any given stage in the algorithm. For example, computing the log-likelihood contribution for an observation occurring from time t_1 to time t_3 might be based on $\bar{\mathbf{y}}_{t_1,t_2}$ and $\bar{\mathbf{y}}_{t_2,t_3}$ rather than directly on $\bar{\mathbf{y}}_{t_1,t_3}$ if another observation began at time t_2.

Even when no instantaneous observations occur at time t, $\mathbf{y}_t$ can be useful to include in the state space if an observation has a time interval beginning or ending at time t. Because of the Markovian property of the Ornstein-Uhlenbeck process, it is possible to compute the distribution of $\bar{\mathbf{y}}_{t_1,t_2}$ from the distribution of $\mathbf{y}_{t_1}$ without need to reference $\bar{\mathbf{y}}_{t_1',t_2'}$ for any

[2]It is possible to make the state space unnecessarily large at some steps to keep the state space constant in dimensionality, but doing so slows likelihood computation and offers no benefit.

Figure A1: Timing for three-observation example



$t'_1 < t'_2 \leq t_1$. As a result, it is convenient to include $\mathbf{y}_t$ in the state space along with values of $\bar{\mathbf{y}}$ covering intervals up to time t. Thus, for each iteration of the algorithm, the state space consists of $\bar{\mathbf{y}}_{t_1,t_2}, \ldots, \bar{\mathbf{y}}_{t_{k-1},t_k}$ and $\mathbf{y}_{t_k}$ for some $t_1 < t_2 < \cdots < t_{k-1} < t_k$.

To illustrate the necessary components of the state space when updating for an observation, consider a simple case with three observations. The first observation is observed from time s_1 to s_3, the second is observed instantaneously at time s_4, and the third is observed from time s_2 to s_5. The timing of the observations is depicted in figure A1. Assume for this example that the latent process is one-dimensional.

To compute the contribution to the log-likelihood from the first observation, $\bar{y}_{s_1,s_2}$ and $\bar{y}_{s_2,s_3}$ are needed, as

$$E\left[\text{obs 1}\right] = \bar{y}_{s_1,s_2}\left(\frac{s_2 - s_1}{s_3 - s_1}\right) + \bar{y}_{s_1,s_2}\left(\frac{s_3 - s_2}{s_3 - s_1}\right). \tag{A4}$$

The state space at this stage also includes y_{s_3}, so as to facilitate future prediction steps, even though observation 1 does not depend upon it. Next, in updating for observation 2, y_{s_4} is needed because $E\left[\text{obs 2}\right] = y\bar{y}_{s_1,s_2}$ and $\bar{y}_{s_2,s_3}$ is needed because it will be needed when updating for observation 3. Finally, observation 3 depends upon $\bar{y}_{s_2,s_3}$, $\bar{y}_{s_3,s_4}$, and $\bar{y}_{s_4,s_5}$.

The prediction stage is more complicated than in many other applications of the Kalman filter. In a theoretical sense, the prediction stage still functions as in the Kalman filter algorithm. However, it is the computation of the appropriate $\mathbf{\Phi}_i$ which is more complicated because it depends on certain results about the behavior of the latent Ornstein-Uhlenbeck process. It is also simpler to perform this update without explicitly computing $\mathbf{\Phi}_i$, although

such a matrix could be produced. The necessary results regarding the latent process appear in section B.

The update stage is less complicated. Minor simplifications relative to the update stage in algorithm A1 are also possible because observations are assumed to be scalar-valued. The one minor addition to this stage is the inclusion of variance transformation parameters γ, ν, and ξ. Algorithm A2 assumes the most general form, although it will often be preferable to assume $\gamma = 0$ or $\xi = 1$ rather than treat all three as free parameters. The update of the log-likelihood is unchanged from algorithm A1.

Finally, the initial values of $\mathbf{x}$ and $\mathbf{V}$ are set to the unconditional distribution of $\mathbf{y}_t$. This mirrors use with ARMA models (Akaike 1978; Anderson and Moore 1979; Gardner, Harvey, and Phillips 1980). This is only possible when the latent process is stationary, as has been assumed in this paper. For the non-stationary case, one alternative is to treat the variance $\mathbf{V}$ as infinite or near-infinite, as has been used for other non-stationary models (Harvey and Phillips 1979; Koopman 1997),[3] although other possible issues exist in a maximum-likelihood framework in this case.

The overall algorithm is given by algorithm A2, although the results in section B are also a necessary component. As an illustration, table A1 gives a basic overview of the steps taken in computing the log-likelihood based on the three-observation example depicted in figure A1 and the state space at each step.

Bayesian estimation An alternative to the use of a maximum-likelihood-based approach is the use of Markov-chain Monte Carlo (MCMC) methods for estimation and inference of MONOCAR models in a Bayesian framework. Some common approaches to Bayesian estimation are inapplicable or ill-suited to MONOCAR. However, efficient Bayesian estimation should be possible through some MCMC methods.

A common approach to Bayesian estimation is the use of software such as WinBUGS,

[3]Other approaches to handle the nonstationary case are also possible, such as computing with $\mathbf{V} = 0$ and adjusting for the uncertainty in the initial value for $\mathbf{x}$ (De Jong 1988).

Algorithm A2: Algorithm for computing the log-likelihood of a MONOCAR model

Data:
- d = the number of latent processes
- N = the number of observations
- **for each** $i = 1$ **to** N
  - z_i = the value of observation i
  - ψ_i^2 = the variance of observation i
  - $g_i \in \{1, \ldots, d\}$ = the latent process measured by observation i
  - $[t_{1,i}, t_{2,i}]$ = time interval for observation i
  - *Note:* $t_{1,i} = t_{2,i}$ `for instantaneous observations`
- **end**

Parameters: Θ, μ, Σ, γ, ν, ξ

Result: ℓ, the log-likelihood of the model

begin
- $\ell \leftarrow 0$
- $S \leftarrow$ the number of unique values of $t_{1,i}$ and $t_{2,i}$
- $s_1, \ldots, s_S \leftarrow$ sorted unique values of $t_{1,i}$ and $t_{2,i}$
- Set the state space to $\mathbf{y}_{s_1}$ with $\mathbf{x} = E\left[\mathbf{y}_{s_1}\right]$ and $\mathbf{V} = Var\left[\mathbf{y}_{s_1}\right]$ (the unconditional mean and variances of $\mathbf{y}_{s_1}$)
- **for** $i = 1$ **to** S **do**
  - **if** $i > 1$ **then**
    - Add $\bar{\mathbf{y}}_{s_{i-1},s_i}$ to the state space with new elements of $\mathbf{x}$ set by $E\left[\bar{\mathbf{y}}_{s_{i-1},s_i}|\mathbf{y}_{s_{i-1}}\right]$ and $\mathbf{V}$ set by $Var\left[\bar{\mathbf{y}}_{s_{i-1},s_i}|\mathbf{y}_{s_{i-1}}\right]$ and $Cov\left[\mathbf{y}_{s_{i-1}}, \bar{\mathbf{y}}_{s_{i-1},s_i}|\mathbf{y}_{s_{i-1}}\right]$
    - Add $\mathbf{y}_{s_i}$ from the state space with new elements of $\mathbf{x}$ set by $E\left[\mathbf{y}_{s_i}|\mathbf{y}_{s_{i-1}}\right]$ and $\mathbf{V}$ set by $Var\left[\mathbf{y}_{s_i}|\mathbf{y}_{s_{i-1}}\right]$, $Cov\left[\mathbf{y}_{s_{i-1}}, \mathbf{y}_{s_i}|\mathbf{y}_{s_{i-1}}\right]$, $Cov\left[\bar{\mathbf{y}}_{s_{i-1},s_i}, \mathbf{y}_{s_i}|\mathbf{y}_{s_{i-1}}\right]$, and $Cov\left[\bar{\mathbf{y}}_{s_{i-1},s_i}, \mathbf{y}_{s_i}|\mathbf{y}_{s_{i-1}}\right]$
    - Remove $\mathbf{y}_{s_{i-1}}$ from the state space, deleting the corresponding elements of $\mathbf{x}$ and $\mathbf{V}$
    - $a \leftarrow a$ **such that** $s_a = \min_{j:t_{2,j} \geq s_i} t_{1,j}$
    - Remove the first components of the state space describing $\bar{\mathbf{y}}_{s_j,s_{j+1}}$ for which $j < a$
    - *Note:* `The state space is now consists of` $\left\{\bar{\mathbf{y}}_{s_a,s_{a+1}}, \ldots, \bar{\mathbf{y}}_{s_{i-1},s_i}, \mathbf{y}_{s_i}\right\}$ `and is` $d(i-a)$ `dimensional`
  - **end**
  - **for each** j **such that** $t_{2,j} = s_i$ **do**
    - $\mathbf{h} \leftarrow$ the $d(i-a)$-element zero vector
    - **if** $t_{1,j} = t_{2,j}$ **then**
      - $\mathbf{h}_{d(i-a)+g_i} \leftarrow 1$
    - **else**
      - *Note:* $t_{1,j} < t_{2,j}$
      - $b \leftarrow b$ **such that** $s_b = t_{1,j}$
      - **for** $k = b + 1$ **to** i **do**
        - $h_{d(k-a)+g_j} = \frac{s_k - s_{k-1}}{s_k - s_b}$
      - **end**
    - **end**
    - $\phi^2 \leftarrow \mathbf{h}^\top \mathbf{V} \mathbf{h} + \gamma_{g_i} + (\nu_{g_i})(\psi_j)^{\xi_{g_i}}$
    - $\ell \leftarrow \ell - \frac{\left(z_i - \mathbf{h}^\top \mathbf{x}\right)^2}{2\phi^2} - \frac{1}{2}\log\left(2\pi\phi^2\right)$
    - $\mathbf{x} \leftarrow \mathbf{x} + \left(\frac{z_i - \mathbf{h}^\top \mathbf{x}}{\phi^2}\right)\mathbf{V}\mathbf{h}$
    - $\mathbf{V} \leftarrow \mathbf{V} - \frac{1}{\phi^2}\mathbf{V}\mathbf{h}\mathbf{h}^\top\mathbf{V}$
  - **end**
- **end**
- **return** ℓ

end

Table A1: Step summary for three-observation example

Action	*State space*
Initialize $\mathbf{x}$ and $\mathbf{V}$	y_{s_1}
Add $\bar{y}_{s_1,s_2}$ and y_{s_2} to state space	$y_{s_1}, \bar{y}_{s_1,s_2}, y_{s_2}$
Remove y_{s_2} from state space	$\bar{y}_{s_1,s_2}, y_{s_2}$
Add $\bar{y}_{s_2,s_3}$ and y_{s_3} to state space	$\bar{y}_{s_1,s_2}, y_{s_2}, \bar{y}_{s_2,s_3}, y_{s_3}$
Remove y_{s_2} from state space	$\bar{y}_{s_1,s_2}, \bar{y}_{s_2,s_3}, y_{s_3}$
Update $\mathbf{x}$, $\mathbf{V}$, and log-likelihood for observation 1	$\bar{y}_{s_1,s_2}, \bar{y}_{s_2,s_3}, y_{s_3}$
Add $\bar{y}_{s_3,s_4}$ and y_{s_4} to state space	$\bar{y}_{s_2,s_3}, y_{s_3}, \bar{y}_{s_3,s_4}, y_{s_4}$
Remove y_{s_3} from state space	$\bar{y}_{s_2,s_3}, \bar{y}_{s_3,s_4}, y_{s_4}$
Update $\mathbf{x}$, $\mathbf{V}$, and log-likelihood for observation 2	$\bar{y}_{s_2,s_3}, \bar{y}_{s_3,s_4}, y_{s_4}$
Add $\bar{y}_{s_4,s_5}$ and y_{s_5} to state space	$\bar{y}_{s_2,s_3}, \bar{y}_{s_3,s_4}, \bar{y}_{s_4,s_5}, y_{s_5}$
Remove y_{s_4} from state space	$\bar{y}_{s_2,s_3}, \bar{y}_{s_3,s_4}, y_{s_4}, \bar{y}_{s_4,s_5}, y_{s_5}$
Update $\mathbf{x}$, $\mathbf{V}$, and log-likelihood for observation 3	$\bar{y}_{s_2,s_3}, \bar{y}_{s_3,s_4}, \bar{y}_{s_4,s_5}, y_{s_5}$

OpenBUGS, JAGS, or Stan to specify a model and provide automated sampling. These software packages are difficult to apply to MONOCAR because they do not allow models with continuous-time processes or observations based on intervals over continuous-time processes and do not offer operators, such as the matrix exponential and kronecker product, needed to compute the covariance between the latent process at different points in time.

MONOCAR is also not an ideal scenario for the use of Gibbs sampling. In a typical latent-variable model estimated via Gibbs sampling, the latent variables are treated as auxiliary parameters. However, because MONOCAR is a continuous-time model, the latent process is infinite-dimensional.[4]

However, other MCMC methods to Bayesian inference are more attractive for Bayesian estimation of MONOCAR. In particular, Hamiltonian Monte Carlo and its extension, the No U-Turn Sampler (Hoffman and Gelman 2014), are compelling possibilities. In addition to the efficiency offered by these samplers, they can be easily combined with the likelihood computation already used for maximum-likelihood estimation of MONOCAR.

[4]While it is possible to reduce the necessary auxiliary parameters needed for Gibbs sampling, doing so loses the simplicity that is usually offered by this approach and still requires a large number of auxiliary parameters.

A Bayesian approach to MONOCAR has several advantages, such as avoiding the need to rely on asymptotic results for inference and allowing prior information to be included. Many advantages offered by a Bayesian approach to MONOCAR are similar to those discussed with other time-series models (see, e.g., Brandt and Freeman 2006, 2009). Priors may also be particularly useful in dealing with parameters used to specify the transformation of the variances of observations described in section 3.4, as we may wish to impose bounds on these parameters to ensure that all variances are positive, which is more problematic in a maximum-likelihood framework, and we may have strong a priori beliefs about plausible values for these parameters.[5] On the other hand, the maximum-likelihood approach offers estimates to be computed much more quickly, allows inference to be performed in a frequentist framework, and avoids the need to specify prior distributions for parameters.

[5]For example, it may be implausible that the actual sampling error associated with a poll is ten times the sampling error implied by the sample size or that the actual sampling error would increase as the sampling size increases.

B. DERIVATION OF DISTRIBUTIONS RELATED TO THE MULTIVARIATE ORNSTEIN-UHLENBECK PROCESS

We begin by defining $\mathbf{y}_t$ by

$$d\mathbf{y}_t = \boldsymbol{\Theta}\left(\boldsymbol{\mu} - \mathbf{y}_t\right) dt + \boldsymbol{\Sigma}^{\frac{1}{2}} d\mathbf{W}_t \tag{A5}$$

where $\mathbf{y}_t$ is n-dimensional; $\mathbf{W}_t$ is an n-dimensional Wiener process; $\boldsymbol{\mu}$, $\boldsymbol{\Theta}$, and $\boldsymbol{\Sigma}$ are parameters (real-valued matrices of size $n \times 1$, $n \times n$, and $n \times n$, respectively), and $\boldsymbol{\Sigma}^{\frac{1}{2}}$ is a matrix satisfying $\left(\boldsymbol{\Sigma}^{\frac{1}{2}}\right)\left(\boldsymbol{\Sigma}^{\frac{1}{2}}\right)^{\mathrm{T}} = \boldsymbol{\Sigma}$. The parameter space is restricted to cases where $\boldsymbol{\Sigma}$ is positive semidefinite, which is necessary and sufficient for $\boldsymbol{\Sigma}^{\frac{1}{2}}$ to exist, and the real parts of the eigenvalues of $\boldsymbol{\Theta}$ are positive, which is sufficient to ensure stationarity. While $\boldsymbol{\Sigma}^{\frac{1}{2}}$ is generally not unique, the distribution of $\mathbf{y}_t$ does not depend on the choice of $\boldsymbol{\Sigma}^{\frac{1}{2}}$ given $\boldsymbol{\Sigma}$.

Assume throughout that $\mathbf{y}_0$ follows a multivariate normal distribution.

B.1. *Related to* $\mathbf{y}_t$

Multiplying both sides of Equation A5 by $e^{\boldsymbol{\Theta}t}$ and noting that $\boldsymbol{\Theta}$ and $e^{\boldsymbol{\Theta}t}$ commute yields

$$e^{\boldsymbol{\Theta}t} d\mathbf{y}_t = \boldsymbol{\Theta} e^{\boldsymbol{\Theta}t}\left(\boldsymbol{\mu} - \mathbf{y}_t\right) dt + e^{\boldsymbol{\Theta}t}\boldsymbol{\Sigma}^{\frac{1}{2}} d\mathbf{W}_t. \tag{A6}$$

Applying Ito's lemma to $f\left(t, x\right) = e^{\boldsymbol{\Theta}t}\mathbf{y}_t$ gives

$$df\left(t, \mathbf{y}_t\right) = \frac{\partial f}{\partial t}\left(t, \mathbf{y}_t\right) dt + \nabla_{\mathbf{y}_t} f\left(t, \mathbf{y}_t\right) d\mathbf{y}_t + \frac{1}{2} d\mathbf{y}_t^{\mathrm{T}}\left(\nabla^2_{\mathbf{y}_t} f\left(t, \mathbf{y}_t\right)\right) d\mathbf{y}_t \tag{A7}$$

$$= \boldsymbol{\Theta} e^{\boldsymbol{\Theta}t}\mathbf{y}_t dt + e^{\boldsymbol{\Theta}t} d\mathbf{y}_t + 0. \tag{A8}$$

Substituting in for $e^{\boldsymbol{\Theta}t} d\mathbf{y}_t$ from Equation A6 gives

$$df\left(t, \mathbf{y}_t\right) = \boldsymbol{\Theta} e^{\boldsymbol{\Theta}t}\mathbf{y}_t dt + \boldsymbol{\Theta} e^{\boldsymbol{\Theta}t}\left(\boldsymbol{\mu} - \mathbf{y}_t\right) dt + e^{\boldsymbol{\Theta}t}\boldsymbol{\Sigma}^{\frac{1}{2}} d\mathbf{W}_t \tag{A9}$$

$$= \boldsymbol{\Theta} e^{\boldsymbol{\Theta}t}\boldsymbol{\mu} + e^{\boldsymbol{\Theta}t}\boldsymbol{\Sigma}^{\frac{1}{2}} d\mathbf{W}_t. \tag{A10}$$

Integrating gives

$$\int_0^t df\left(s, \mathbf{y}_s\right) = \int_0^t \boldsymbol{\Theta} e^{\boldsymbol{\Theta} s} \boldsymbol{\mu}\, ds + \int_0^t e^{\boldsymbol{\Theta} s} \boldsymbol{\Sigma}^{\frac{1}{2}} d\mathbf{W}_s \tag{A11}$$

$$e^{\boldsymbol{\Theta} t}\mathbf{y}_t - \mathbf{y}_0 = \left(e^{\boldsymbol{\Theta} t} - \mathbf{I}\right)\boldsymbol{\mu} + \int_0^t e^{\boldsymbol{\Theta} s} \boldsymbol{\Sigma}^{\frac{1}{2}} d\mathbf{W}_s \tag{A12}$$

$$\mathbf{y}_t - e^{-\boldsymbol{\Theta} t}\mathbf{y}_0 = \left(\mathbf{I} - e^{-\boldsymbol{\Theta} t}\right)\boldsymbol{\mu} + \int_0^t e^{-\boldsymbol{\Theta}(t-s)} \boldsymbol{\Sigma}^{\frac{1}{2}} d\mathbf{W}_s \tag{A13}$$

$$\mathbf{y}_t = e^{-\boldsymbol{\Theta} t}\mathbf{y}_0 + \left(\mathbf{I} - e^{-\boldsymbol{\Theta} t}\right)\boldsymbol{\mu} + \int_0^t e^{-\boldsymbol{\Theta}(t-s)} \boldsymbol{\Sigma}^{\frac{1}{2}} d\mathbf{W}_s. \tag{A14}$$

Distribution of $\mathbf{y}_t$ *given the distribution of* $\mathbf{y}_0$ Since $\mathbf{y}_0$ follows a multivariate normal distribution, so does the linear transformation $e^{-\boldsymbol{\Theta} t}\mathbf{y}_0$. Since $\left(\mathbf{I} - e^{-\boldsymbol{\Theta} t}\right)\boldsymbol{\mu}$ is a constant and $\int_0^t e^{-\boldsymbol{\Theta}(t-s)} \boldsymbol{\Sigma}^{\frac{1}{2}} d\mathbf{W}_s$ is a linear operator on a Wiener process independent of $\mathbf{y}_0$—essentially a sum of infinitely many independent multivariate normal distributions—$\mathbf{y}_t$ must also follow a multivariate normal distribution, $\mathbf{y}_t$ must therefore follow a multivariate normal distribution.

We can compute the expectation of $\mathbf{y}_t$ as

$$E\left[\mathbf{y}_t\right] = E\left[e^{-\boldsymbol{\Theta} t}\mathbf{y}_0\right] + \left(\mathbf{I} - e^{-\boldsymbol{\Theta} t}\right)\boldsymbol{\mu} + E\left[\int_0^t e^{-\boldsymbol{\Theta}(t-s)} \boldsymbol{\Sigma}^{\frac{1}{2}} d\mathbf{W}_s\right] \tag{A15}$$

$$= e^{-\boldsymbol{\Theta} t} E\left[\mathbf{y}_0\right] + \left(\mathbf{I} - e^{-\boldsymbol{\Theta} t}\right)\boldsymbol{\mu} + \int_0^t e^{-\boldsymbol{\Theta}(t-s)} \boldsymbol{\Sigma}^{\frac{1}{2}} E\left[d\mathbf{W}_s\right] \tag{A16}$$

$$= e^{-\boldsymbol{\Theta} t} E\left[\mathbf{y}_0\right] + \left(\mathbf{I} - e^{-\boldsymbol{\Theta} t}\right)\boldsymbol{\mu} \tag{A17}$$

and the variance as

$$Var\left[\mathbf{y}_t\right] = Var\left[e^{-\boldsymbol{\Theta} t}\mathbf{y}_0\right] + Var\left[\left(\mathbf{I} - e^{-\boldsymbol{\Theta} t}\right)\boldsymbol{\mu}\right] + Var\left[\int_0^t e^{-\boldsymbol{\Theta}(t-s)} \boldsymbol{\Sigma}^{\frac{1}{2}} d\mathbf{W}_s\right] \tag{A18}$$

$$= e^{-\boldsymbol{\Theta} t}\, Var\left[\mathbf{y}_0\right] e^{-\boldsymbol{\Theta}^{\mathrm{T}} t} + Var\left[e^{-\boldsymbol{\Theta} t}\mathbf{y}_0\right]. \tag{A19}$$

Since $E\left[\int_0^t e^{-\boldsymbol{\Theta}(t-s)} \boldsymbol{\Sigma}^{\frac{1}{2}} d\mathbf{W}_s\right] = 0$, $d\mathbf{W}_s d\mathbf{W}_s^{\mathrm{T}} = I ds$, and $d\mathbf{W}_s d\mathbf{W}_u^{\mathrm{T}} = 0$ for $s \neq u$,

$$Var\left[\int_0^t e^{-\boldsymbol{\Theta}(t-s)} \boldsymbol{\Sigma}^{\frac{1}{2}} d\mathbf{W}_s\right] = \int_0^t \int_0^t \left\{e^{-\boldsymbol{\Theta}(t-s)} \boldsymbol{\Sigma}^{\frac{1}{2}} d\mathbf{W}_s ds \mathbf{W}_u^{\mathrm{T}} \left(\boldsymbol{\Sigma}^{\frac{1}{2}}\right)^{\mathrm{T}} e^{-\boldsymbol{\Theta}^{\mathrm{T}}(t-u)}\right\} \tag{A20}$$

$$= \int_0^t \left\{e^{-\boldsymbol{\Theta}(t-s)} \boldsymbol{\Sigma} e^{-\boldsymbol{\Theta}^{\mathrm{T}}(t-s)}\right\} ds. \tag{A21}$$

Thus, the identity $\text{vec}(ABC) = \left(C^{\mathrm{T}} \otimes A\right) \text{vec}(B)$ gives us

$$\text{vec}\left(Var\left[\int_0^t e^{-\boldsymbol{\Theta}(t-s)} \boldsymbol{\Sigma}^{\frac{1}{2}} d\mathbf{W}_s \right] \right) = \int_0^t \left(e^{-\boldsymbol{\Theta}(t-s)} \otimes e^{-\boldsymbol{\Theta}(t-s)} \right) \text{vec}(\boldsymbol{\Sigma})\, ds \tag{A22}$$

$$= \int_0^t \left(e^{-(\boldsymbol{\Theta} \oplus \boldsymbol{\Theta})(t-s)} \right) \text{vec}(\boldsymbol{\Sigma})\, ds \tag{A23}$$

$$= (\boldsymbol{\Theta} \oplus \boldsymbol{\Theta})^{-1} \left(\mathbf{I} - e^{-(\boldsymbol{\Theta} \oplus \boldsymbol{\Theta})t} \right) \text{vec}(\boldsymbol{\Sigma}) \tag{A24}$$

So, $\mathbf{y}_t$ follows a multivariate normal distribution with

$$E[\mathbf{y}_t] = e^{-\boldsymbol{\Theta} t} E[\mathbf{y}_0] + \left(\mathbf{I} - e^{-\boldsymbol{\Theta} t} \right) \boldsymbol{\mu} \tag{A25}$$

and

$$\text{vec}(\, Var[\mathbf{y}_t]) = \text{vec}\left(e^{-\boldsymbol{\Theta} t}\, Var[\mathbf{y}_0]\, e^{-\boldsymbol{\Theta}^{\mathrm{T}} t} \right) + (\boldsymbol{\Theta} \oplus \boldsymbol{\Theta})^{-1} \left(\mathbf{I} - e^{-(\boldsymbol{\Theta} \oplus \boldsymbol{\Theta})t} \right) \text{vec}(\boldsymbol{\Sigma}) \tag{A26}$$

$$= e^{-(\boldsymbol{\Theta} \oplus \boldsymbol{\Theta})t} \text{vec}(\, Var[\mathbf{y}_0]) + (\boldsymbol{\Theta} \oplus \boldsymbol{\Theta})^{-1} \left(\mathbf{I} - e^{-(\boldsymbol{\Theta} \oplus \boldsymbol{\Theta})t} \right) \text{vec}(\boldsymbol{\Sigma}) \tag{A27}$$

B.2. *Related to* $\bar{\mathbf{y}}_{0,t}$

Let $\bar{\mathbf{y}}_{0,t}$ represent the arithmetic mean of $\mathbf{y}_t$ over the interval $[0, t]$. Thus,

$$t\bar{\mathbf{y}}_{0,t} = \int_0^t \mathbf{y}_s ds \tag{A28}$$

$$= \int_0^t e^{-\boldsymbol{\Theta} s} \mathbf{y}_0 ds + \int_0^t \left(\mathbf{I} - e^{-\boldsymbol{\Theta} s} \right) \boldsymbol{\mu}\, ds + \int_0^t \int_0^s e^{-\boldsymbol{\Theta}(s-u)} \boldsymbol{\Sigma}^{\frac{1}{2}} d\mathbf{W}_u ds \tag{A29}$$

$$= -\boldsymbol{\Theta}^{-1} \left(e^{-\boldsymbol{\Theta} t} - \mathbf{I} \right) \mathbf{y}_0 + \left(t\mathbf{I} + \boldsymbol{\Theta}^{-1} \left(e^{-\boldsymbol{\Theta} t} - \mathbf{I} \right) \right) \boldsymbol{\mu} + \int_0^t \int_0^s e^{-\boldsymbol{\Theta}(s-u)} \boldsymbol{\Sigma}^{\frac{1}{2}} d\mathbf{W}_u ds \tag{A30}$$

$$= \boldsymbol{\Theta}^{-1} \left(e^{-\boldsymbol{\Theta} t} - \mathbf{I} \right) (\boldsymbol{\mu} - \mathbf{y}_0) + t\boldsymbol{\mu} + \int_0^t \int_0^s e^{-\boldsymbol{\Theta}(s-u)} \boldsymbol{\Sigma}^{\frac{1}{2}} d\mathbf{W}_u ds \tag{A31}$$

$$\bar{\mathbf{y}}_{0,t} = \frac{1}{t} \boldsymbol{\Theta}^{-1} \left(e^{-\boldsymbol{\Theta} t} - \mathbf{I} \right) (\boldsymbol{\mu} - \mathbf{y}_0) + \boldsymbol{\mu} + \frac{1}{t} \int_0^t \int_0^s e^{-\boldsymbol{\Theta}(s-u)} \boldsymbol{\Sigma}^{\frac{1}{2}} d\mathbf{W}_u ds \tag{A32}$$

Distribution of $\bar{\mathbf{y}}_{0,t}$ *given the distribution of* $\mathbf{y}_0$ Since $\mathbf{y}_0$ follows a multivariate normal distribution, $\frac{1}{t} \boldsymbol{\Theta}^{-1} \left(e^{-\boldsymbol{\Theta} t} - \mathbf{I} \right) (\boldsymbol{\mu} - \mathbf{y}_0)$ must also follow a multivariate normal distribution as it is an affine transformation of $\mathbf{y}_0$. Note that $\int_0^s e^{-\boldsymbol{\Theta}(s-u)} \boldsymbol{\Sigma}^{\frac{1}{2}} d\mathbf{W}_u$ must follow a multivariate normal distribution independent of $\mathbf{y}_0$ as before, and, thus, so must $\frac{1}{t} \int_0^t \int_0^s e^{-\boldsymbol{\Theta}(s-u)} \boldsymbol{\Sigma}^{\frac{1}{2}} d\mathbf{W}_u ds$. Since $\boldsymbol{\mu}$ is a constant, we have that $\bar{\mathbf{y}}_{0,t}$ is a sum of independent multivariate normal random variables and must, therefore, itself follow a multivariate normal distribution.

We can compute the expectation of $\bar{\mathbf{y}}_{0,t}$ as

$$E\left[\bar{\mathbf{y}}_{0,t}\right] = E\left[\frac{1}{t}\boldsymbol{\Theta}^{-1}\left(e^{-\boldsymbol{\Theta}t} - \mathbf{I}\right)\left(\boldsymbol{\mu} - \mathbf{y}_0\right)\right] + \boldsymbol{\mu} + E\left[\frac{1}{t}\int_0^t \int_0^s e^{-\boldsymbol{\Theta}(s-u)}\boldsymbol{\Sigma}^{\frac{1}{2}} d\mathbf{W}_u ds\right] \tag{A33}$$

$$= \frac{1}{t}\boldsymbol{\Theta}^{-1}\left(e^{-\boldsymbol{\Theta}t} - \mathbf{I}\right)\left(\boldsymbol{\mu} - E\left[\mathbf{y}_0\right]\right) + \boldsymbol{\mu} + \frac{1}{t}\int_0^t \int_0^s e^{-\boldsymbol{\Theta}(s-u)}\boldsymbol{\Sigma}^{\frac{1}{2}} E\left[d\mathbf{W}_u\right] ds \tag{A34}$$

$$= \frac{1}{t}\boldsymbol{\Theta}^{-1}\left(e^{-\boldsymbol{\Theta}t} - \mathbf{I}\right)\left(\boldsymbol{\mu} - E\left[\mathbf{y}_0\right]\right) + \boldsymbol{\mu} \tag{A35}$$

and the variance as

$$\begin{aligned} Var\left[\bar{\mathbf{y}}_{0,t}\right] = {} & Var\left[\frac{1}{t}\boldsymbol{\Theta}^{-1}\left(e^{-\boldsymbol{\Theta}t} - \mathbf{I}\right)\left(\boldsymbol{\mu} - \mathbf{y}_0\right)\right] + Var\left[\boldsymbol{\mu}\right] \\ & + Var\left[\frac{1}{t}\int_0^t \int_0^s e^{-\boldsymbol{\Theta}(s-u)}\boldsymbol{\Sigma}^{\frac{1}{2}} d\mathbf{W}_u ds\right] \end{aligned} \tag{A36}$$

$$\begin{aligned} = {} & \frac{1}{t^2}\boldsymbol{\Theta}^{-1}\left(e^{-\boldsymbol{\Theta}t} - \mathbf{I}\right) Var\left[\mathbf{y}_0\right]\left(e^{-\boldsymbol{\Theta}^{\mathrm{T}}t} - \mathbf{I}\right)\left(\boldsymbol{\Theta}^{-1}\right)^{\mathrm{T}} \\ & + \frac{1}{t^2} Var\left[\int_0^t \int_0^s e^{-\boldsymbol{\Theta}(s-u)}\boldsymbol{\Sigma}^{\frac{1}{2}} d\mathbf{W}_u ds\right]. \end{aligned} \tag{A37}$$

Since $E\left[\int_0^t \int_0^s e^{-\boldsymbol{\Theta}(s-u)}\boldsymbol{\Sigma}^{\frac{1}{2}} d\mathbf{W}_u ds\right] = 0$,

$$Var\left[\int_0^t \int_0^s e^{-\boldsymbol{\Theta}(s-u)}\boldsymbol{\Sigma}^{\frac{1}{2}} d\mathbf{W}_u ds\right] = E\left[\int_0^t \int_0^s e^{-\boldsymbol{\Theta}(s-u)}\boldsymbol{\Sigma}^{\frac{1}{2}} d\mathbf{W}_u ds \int_0^t \int_0^z d\mathbf{W}_v^{\mathrm{T}}\left(\boldsymbol{\Sigma}^{\frac{1}{2}}\right)^{\mathrm{T}} e^{-\boldsymbol{\Theta}^{\mathrm{T}}(z-v)} dz\right] \tag{A38}$$

$$= E\left[\int_0^t \int_0^t \int_0^z \int_0^s e^{-\boldsymbol{\Theta}(s-u)}\boldsymbol{\Sigma}^{\frac{1}{2}} d\mathbf{W}_u d\mathbf{W}_v^{\mathrm{T}}\left(\boldsymbol{\Sigma}^{\frac{1}{2}}\right)^{\mathrm{T}} e^{-\boldsymbol{\Theta}^{\mathrm{T}}(z-v)} dz\, ds\right] \tag{A39}$$

$$= \int_0^t \int_0^t \int_0^z \int_0^s e^{-\boldsymbol{\Theta}(s-u)} E\left[\boldsymbol{\Sigma}^{\frac{1}{2}} d\mathbf{W}_u d\mathbf{W}_v^{\mathrm{T}}\left(\boldsymbol{\Sigma}^{\frac{1}{2}}\right)^{\mathrm{T}}\right] e^{-\boldsymbol{\Theta}^{\mathrm{T}}(z-v)} dz\, ds \tag{A40}$$

$$= \int_0^t \int_0^t \int_0^{\min(s,z)} e^{-\boldsymbol{\Theta}(s-u)}\boldsymbol{\Sigma}^{\frac{1}{2}} I du\left(\boldsymbol{\Sigma}^{\frac{1}{2}}\right)^{\mathrm{T}} e^{-\boldsymbol{\Theta}^{\mathrm{T}}(z-u)} dz\, ds \tag{A41}$$

$$= \int_0^t \int_0^t \int_0^{\min(s,z)} e^{-\boldsymbol{\Theta}(s-u)}\boldsymbol{\Sigma} e^{-\boldsymbol{\Theta}^{\mathrm{T}}(z-u)} du\, dz\, ds. \tag{A42}$$

Thus,

$$\operatorname{vec}\left(Var\left[\int_0^t \int_0^s e^{-\boldsymbol{\Theta}(t-u)} \boldsymbol{\Sigma}^{\frac{1}{2}} d\mathbf{W}_u ds\right]\right) = \operatorname{vec}\left(\int_0^t \int_0^t \int_0^{\min(s,z)} e^{-\boldsymbol{\Theta}(s-u)} \boldsymbol{\Sigma} e^{-\boldsymbol{\Theta}^{\mathrm{T}}(z-u)} du\, dz\, ds\right) \tag{A43}$$

$$= \int_0^t \int_0^t \int_0^{\min(s,z)} \operatorname{vec}\left(e^{-\boldsymbol{\Theta}(s-u)} \boldsymbol{\Sigma} e^{-\boldsymbol{\Theta}^{\mathrm{T}}(z-u)}\right) du\, dz\, ds \tag{A44}$$

$$= \int_0^t \int_0^s \int_0^z \operatorname{vec}\left(e^{-\boldsymbol{\Theta}(s-u)} \boldsymbol{\Sigma} e^{-\boldsymbol{\Theta}^{\mathrm{T}}(z-u)}\right) du\, dz\, ds \tag{A45}$$

$$+ \int_0^t \int_0^z \int_0^s \operatorname{vec}\left(e^{-\boldsymbol{\Theta}(s-u)} \boldsymbol{\Sigma} e^{-\boldsymbol{\Theta}^{\mathrm{T}}(z-u)}\right) du\, ds\, dz \tag{A46}$$

Since $\boldsymbol{\Theta}$, $e^{-\boldsymbol{\Theta}(s-z)}$, and I all commute, $(\boldsymbol{\Theta} \oplus \boldsymbol{\Theta})^{-1}$ and $\left(e^{-\boldsymbol{\Theta}(s-z)} \otimes \mathbf{I}\right)$ must commute. Thus, we can expand the first half of the above representation to yield

$$\int_0^t \int_0^s \int_0^z \operatorname{vec}\left(e^{-\boldsymbol{\Theta}(s-u)} \boldsymbol{\Sigma} e^{-\boldsymbol{\Theta}^{\mathrm{T}}(z-u)}\right) du\, dz\, ds \tag{A47}$$

$$= \int_0^t \int_0^s \int_0^z \left(e^{-\boldsymbol{\Theta}(s-u)} \otimes e^{-\boldsymbol{\Theta}(z-u)}\right) \operatorname{vec}(\boldsymbol{\Sigma})\, du\, dz\, ds \tag{A48}$$

$$= \int_0^t \int_0^s \int_0^z \left(e^{-\boldsymbol{\Theta}(s-z)} e^{-\boldsymbol{\Theta}(z-u)} \otimes e^{-\boldsymbol{\Theta}(z-u)}\right) du\, dz\, ds \operatorname{vec}(\boldsymbol{\Sigma}) \tag{A49}$$

$$= \int_0^t \int_0^s \int_0^z \left(e^{-\boldsymbol{\Theta}(s-z)} \otimes \mathbf{I}\right) \left(e^{-\boldsymbol{\Theta}(z-u)} \otimes e^{-\boldsymbol{\Theta}(z-u)}\right) du\, dz\, ds \operatorname{vec}(\boldsymbol{\Sigma}) \tag{A50}$$

$$= \int_0^t \int_0^s \int_0^z \left(e^{-\boldsymbol{\Theta}(s-z)} \otimes \mathbf{I}\right) \left(e^{-(\boldsymbol{\Theta}\oplus\boldsymbol{\Theta})(z-u)}\right) du\, dz\, ds \operatorname{vec}(\boldsymbol{\Sigma}) \tag{A51}$$

$$= \int_0^t \int_0^s \left(e^{-\boldsymbol{\Theta}(s-z)} \otimes \mathbf{I}\right) \int_0^z e^{-(\boldsymbol{\Theta}\oplus\boldsymbol{\Theta})(z-u)} du\, dz\, ds \operatorname{vec}(\boldsymbol{\Sigma}) \tag{A52}$$

$$= \int_0^t \int_0^s \left(e^{-\boldsymbol{\Theta}(s-z)} \otimes \mathbf{I}\right) (\boldsymbol{\Theta} \oplus \boldsymbol{\Theta})^{-1} \left(\mathbf{I} - e^{-(\boldsymbol{\Theta}\oplus\boldsymbol{\Theta})z}\right) dz\, ds \operatorname{vec}(\boldsymbol{\Sigma}) \tag{A53}$$

$$= (\boldsymbol{\Theta} \oplus \boldsymbol{\Theta})^{-1} \int_0^t \int_0^s \left(e^{-\boldsymbol{\Theta}(s-z)} \otimes \mathbf{I}\right) \left(\mathbf{I} - e^{-(\boldsymbol{\Theta}\oplus\boldsymbol{\Theta})z}\right) dz\, ds \operatorname{vec}(\boldsymbol{\Sigma}). \tag{A54}$$

Looking only at the inner part of the right side of Equation A54, we have

$$\int_0^t \int_0^s \left(e^{-\mathbf{\Theta}(s-z)} \otimes \mathbf{I}\right) \left(\mathbf{I} - e^{-(\mathbf{\Theta}\oplus\mathbf{\Theta})z}\right) dz\, ds \tag{A55}$$

$$= \int_0^t \int_0^s \left(e^{-\mathbf{\Theta}(s-z)} \otimes \mathbf{I}\right) dz\, ds - \int_0^t \int_0^s \left(e^{-\mathbf{\Theta}(s-z)} \otimes \mathbf{I}\right) \left(e^{-\mathbf{\Theta}z} \otimes e^{-\mathbf{\Theta}z}\right) dz\, ds \tag{A56}$$

$$= \int_0^t \int_0^s \left(e^{-\mathbf{\Theta}(s-z)} \otimes \mathbf{I}\right) dz\, ds - \int_0^t \int_0^s \left(e^{-\mathbf{\Theta}s} \otimes e^{-\mathbf{\Theta}z}\right) dz\, ds \tag{A57}$$

$$= \int_0^t \int_0^s e^{-\mathbf{\Theta}(s-z)} dz\, ds \otimes \mathbf{I} - \int_0^t \left(e^{-\mathbf{\Theta}s} \otimes \int_0^s e^{-\mathbf{\Theta}z} dz\right) ds \tag{A58}$$

$$= \int_0^t \mathbf{\Theta}^{-1} \left(\mathbf{I} - e^{-\mathbf{\Theta}s}\right) ds \otimes \mathbf{I} - \int_0^t \left(e^{-\mathbf{\Theta}s} \otimes \left(-\mathbf{\Theta}^{-1}\right) \left(e^{-\mathbf{\Theta}s} - \mathbf{I}\right)\right) ds \tag{A59}$$

$$= \mathbf{\Theta}^{-1} \left(t\mathbf{I} + \mathbf{\Theta}^{-1} \left(e^{-\mathbf{\Theta}t} - \mathbf{I}\right)\right) \otimes \mathbf{I} + \int_0^t \left(\mathbf{I} \otimes \mathbf{\Theta}^{-1}\right) \left(e^{-\mathbf{\Theta}s} \otimes \left(e^{-\mathbf{\Theta}s} - \mathbf{I}\right)\right) ds \tag{A60}$$

$$\begin{aligned} &= \left(\mathbf{\Theta}^{-1} \left(t\mathbf{I} + \mathbf{\Theta}^{-1} \left(e^{-\mathbf{\Theta}t} - \mathbf{I}\right)\right) \otimes \mathbf{I}\right) \\ &\quad + \left(\mathbf{I} \otimes \mathbf{\Theta}^{-1}\right) \int_0^t \left(e^{-(\mathbf{\Theta}\oplus\mathbf{\Theta})s}\right) ds - \left(\mathbf{I} \otimes \mathbf{\Theta}^{-1}\right) \int_0^t \left(e^{-\mathbf{\Theta}s} \otimes \mathbf{I}\right) ds \end{aligned} \tag{A61}$$

$$\begin{aligned} &= \left(\mathbf{\Theta}^{-1} \left(t\mathbf{I} + \mathbf{\Theta}^{-1} \left(e^{-\mathbf{\Theta}t} - \mathbf{I}\right)\right) \otimes \mathbf{I}\right) \\ &\quad + \left(\mathbf{I} \otimes \mathbf{\Theta}^{-1}\right) \left(-\left(\mathbf{\Theta} \oplus \mathbf{\Theta}\right)^{-1}\right) \left(e^{-(\mathbf{\Theta}\oplus\mathbf{\Theta})t} - \mathbf{I}\right) - \left(\mathbf{I} \otimes \mathbf{\Theta}^{-1}\right) \left(\int_0^t e^{-\mathbf{\Theta}s} ds \otimes \mathbf{I}\right) \end{aligned} \tag{A62}$$

$$\begin{aligned} &= \left(\mathbf{\Theta}^{-1} \left(t\mathbf{I} + \mathbf{\Theta}^{-1} \left(e^{-\mathbf{\Theta}t} - \mathbf{I}\right)\right) \otimes \mathbf{I}\right) \\ &\quad - \left(\mathbf{I} \otimes \mathbf{\Theta}^{-1}\right) \left(\mathbf{\Theta} \oplus \mathbf{\Theta}\right)^{-1} \left(e^{-(\mathbf{\Theta}\oplus\mathbf{\Theta})t} - \mathbf{I}\right) + \left(\mathbf{I} \otimes \mathbf{\Theta}^{-1}\right) \left(\mathbf{\Theta}^{-1} \left(e^{-\mathbf{\Theta}t} - \mathbf{I}\right) \otimes \mathbf{I}\right) \end{aligned} \tag{A63}$$

$$\begin{aligned} &= \left(\mathbf{\Theta}^{-1} \left(t\mathbf{I} + \mathbf{\Theta}^{-1} \left(e^{-\mathbf{\Theta}t} - \mathbf{I}\right)\right) \otimes \mathbf{I}\right) \\ &\quad - \left(\mathbf{I} \otimes \mathbf{\Theta}^{-1}\right) \left(\mathbf{\Theta} \oplus \mathbf{\Theta}\right)^{-1} \left(e^{-(\mathbf{\Theta}\oplus\mathbf{\Theta})t} - \mathbf{I}\right) + \left(\mathbf{\Theta}^{-1} \left(e^{-\mathbf{\Theta}t} - \mathbf{I}\right) \otimes \mathbf{\Theta}^{-1}\right). \end{aligned} \tag{A64}$$

$$\tag{A65}$$

It will be useful to refer to this quantity as $\mathbf{\Psi}$. Thus, we are left with

$$\int_0^t \int_0^s \int_0^z \mathrm{vec}\left(e^{-\mathbf{\Theta}(s-u)} \mathbf{\Sigma} e^{-\mathbf{\Theta}^{\mathrm{T}}(z-u)}\right) du\, dz\, ds = \left(\mathbf{\Theta} \oplus \mathbf{\Theta}\right)^{-1} \mathbf{\Psi} \mathrm{vec}\left(\mathbf{\Sigma}\right) \tag{A66}$$

where

$$\begin{aligned} \mathbf{\Psi} = &\mathbf{\Theta}^{-1} \left(t\mathbf{I} + \mathbf{\Theta}^{-1} \left(e^{-\mathbf{\Theta}t} - \mathbf{I}\right)\right) \otimes \mathbf{I} + \mathbf{\Theta}^{-1} \left(e^{-\mathbf{\Theta}t} - \mathbf{I}\right) \otimes \mathbf{\Theta}^{-1} \\ &+ \left(\mathbf{I} \otimes \mathbf{\Theta}^{-1}\right) \left(\mathbf{\Theta} \oplus \mathbf{\Theta}\right)^{-1} - \left(\mathbf{I} \otimes \mathbf{\Theta}^{-1}\right) \left(\mathbf{\Theta} \oplus \mathbf{\Theta}\right)^{-1} e^{-(\mathbf{\Theta}\oplus\mathbf{\Theta})t}. \end{aligned} \tag{A67}$$

By a similar derivation,

$$\int_0^t \int_0^z \int_0^s \mathrm{vec}\left(e^{-\boldsymbol{\Theta}(s-u)}\boldsymbol{\Sigma}e^{-\boldsymbol{\Theta}^{\mathrm{T}}(z-u)}\right) du\, ds\, dz = (\boldsymbol{\Theta}\oplus\boldsymbol{\Theta})^{-1}\boldsymbol{\Xi}\mathrm{vec}\left(\boldsymbol{\Sigma}\right). \tag{A68}$$

where

$$\begin{aligned}\boldsymbol{\Xi} = \mathbf{I}\otimes\boldsymbol{\Theta}^{-1}\left(t\mathbf{I}+\boldsymbol{\Theta}^{-1}\left(e^{-\boldsymbol{\Theta}t}-\mathbf{I}\right)\right) + \boldsymbol{\Theta}^{-1}\otimes\boldsymbol{\Theta}^{-1}\left(e^{-\boldsymbol{\Theta}t}-\mathbf{I}\right)\\ + \left(\mathbf{I}\otimes\boldsymbol{\Theta}^{-1}\right)(\boldsymbol{\Theta}\oplus\boldsymbol{\Theta})^{-1} - \left(\mathbf{I}\otimes\boldsymbol{\Theta}^{-1}\right)(\boldsymbol{\Theta}\oplus\boldsymbol{\Theta})^{-1}e^{-(\boldsymbol{\Theta}\oplus\boldsymbol{\Theta})t}\end{aligned} \tag{A69}$$

Let $\boldsymbol{\Upsilon}$ be the matrix such that

$$\mathrm{vec}\left(\boldsymbol{\Upsilon}\right) = \int_0^t \int_0^s \int_0^z \mathrm{vec}\left(e^{-\boldsymbol{\Theta}(s-u)}\boldsymbol{\Sigma}e^{-\boldsymbol{\Theta}^{\mathrm{T}}(z-u)}\right) du\, dz\, ds \tag{A70}$$

$$= (\boldsymbol{\Theta}\oplus\boldsymbol{\Theta})^{-1}\boldsymbol{\Psi}\mathrm{vec}\left(\boldsymbol{\Sigma}\right). \tag{A71}$$

Thus,

$$\mathrm{vec}\left(\boldsymbol{\Upsilon}^{\mathrm{T}}\right) = \int_0^t \int_0^s \int_0^z \mathrm{vec}\left(e^{-\boldsymbol{\Theta}(z-u)}\boldsymbol{\Sigma}e^{-\boldsymbol{\Theta}^{\mathrm{T}}(s-u)}\right) du\, dz\, ds \tag{A72}$$

$$= \int_0^t \int_0^z \int_0^s \mathrm{vec}\left(e^{-\boldsymbol{\Theta}(s-u)}\boldsymbol{\Sigma}e^{-\boldsymbol{\Theta}^{\mathrm{T}}(z-u)}\right) du\, ds\, dz \tag{A73}$$

$$= (\boldsymbol{\Theta}\oplus\boldsymbol{\Theta})^{-1}\boldsymbol{\Xi}\mathrm{vec}\left(\boldsymbol{\Sigma}\right). \tag{A74}$$

It follows that

$$\begin{aligned}\mathrm{vec}\left(Var\left[\int_0^t\int_0^s e^{-\boldsymbol{\Theta}(t-u)}\boldsymbol{\Sigma}^{\frac{1}{2}}d\mathbf{W}_u ds\right]\right) &= \int_0^t \int_0^s \int_0^z \mathrm{vec}\left(e^{-\boldsymbol{\Theta}(s-u)}\boldsymbol{\Sigma}e^{-\boldsymbol{\Theta}^{\mathrm{T}}(z-u)}\right) du\, dz\, ds\\ &\quad + \int_0^t \int_0^z \int_0^s \mathrm{vec}\left(e^{-\boldsymbol{\Theta}(s-u)}\boldsymbol{\Sigma}e^{-\boldsymbol{\Theta}^{\mathrm{T}}(z-u)}\right) du\, ds\, dz\end{aligned} \tag{A75}$$

$$= \mathrm{vec}\left(\boldsymbol{\Upsilon}\right) + \mathrm{vec}\left(\boldsymbol{\Upsilon}^{\mathrm{T}}\right) \tag{A76}$$

$$= \mathrm{vec}\left(\boldsymbol{\Upsilon}+\boldsymbol{\Upsilon}^{\mathrm{T}}\right) \tag{A77}$$

and, so,

$$Var\left[\int_0^t\int_0^s e^{-\boldsymbol{\Theta}(t-u)}\boldsymbol{\Sigma}^{\frac{1}{2}}d\mathbf{W}_u ds\right] = \boldsymbol{\Upsilon}+\boldsymbol{\Upsilon}^{\mathrm{T}}. \tag{A78}$$

Therefore, $\bar{\mathbf{y}}_{0,t}$ follows a multivariate normal distribution with

$$E\left[\bar{\mathbf{y}}_{0,t}\right] = \frac{1}{t}\boldsymbol{\Theta}^{-1}\left(e^{-\boldsymbol{\Theta}t} - \mathbf{I}\right)\left(\boldsymbol{\mu} - E\left[\mathbf{y}_0\right]\right) + \boldsymbol{\mu} \tag{A79}$$

and

$$\begin{aligned} Var\left[\bar{\mathbf{y}}_{0,t}\right] &= \frac{1}{t^2}\boldsymbol{\Theta}^{-1}\left(e^{-\boldsymbol{\Theta}t} - \mathbf{I}\right) Var\left[\mathbf{y}_0\right]\left(e^{-\boldsymbol{\Theta}^{\mathrm{T}}t} - \mathbf{I}\right)\left(\boldsymbol{\Theta}^{-1}\right)^{\mathrm{T}} \\ &\quad + Var\left[\int_0^t \int_0^s e^{-\boldsymbol{\Theta}(t-u)}\boldsymbol{\Sigma}^{\frac{1}{2}} d\mathbf{W}_u ds\right] \end{aligned} \tag{A80}$$

$$= \frac{1}{t^2}\boldsymbol{\Theta}^{-1}\left(e^{-\boldsymbol{\Theta}t} - \mathbf{I}\right) Var\left[\mathbf{y}_0\right]\left(e^{-\boldsymbol{\Theta}^{\mathrm{T}}t} - \mathbf{I}\right)\left(\boldsymbol{\Theta}^{-1}\right)^{\mathrm{T}} + \boldsymbol{\Upsilon} + \boldsymbol{\Upsilon}^{\mathrm{T}} \tag{A81}$$

where $\boldsymbol{\Upsilon}$ is defined in Equation A71.

B.3. *Covariances*

Note that $\bar{\mathbf{y}}_{0,t}$ and $\mathbf{y}_t$ are formed from a linear operator applied to a Wiener process and $\mathbf{y}_0$, which is itself multivariate normal. Therefore, the joint distribution of $\left(\mathbf{y}_0^{\mathrm{T}}, \bar{\mathbf{y}}_{0,t}^{\mathrm{T}}, \mathbf{y}_t^{\mathrm{T}}\right)^{\mathrm{T}}$ also follows a multivariate normal distribution.

Covariance between $\mathbf{y}_0$ *and* $\bar{\mathbf{y}}_{0,t}$ Since

$$\bar{\mathbf{y}}_{0,t} = \frac{1}{t}\boldsymbol{\Theta}^{-1}\left(e^{-\boldsymbol{\Theta}t} - \mathbf{I}\right)\left(\boldsymbol{\mu} - \mathbf{y}_0\right) + \boldsymbol{\mu} + \frac{1}{t}\int_0^t \int_0^s e^{-\boldsymbol{\Theta}(t-u)}\boldsymbol{\Sigma}^{\frac{1}{2}} d\mathbf{W}_u ds \tag{A82}$$

and $\frac{1}{t}\int_0^t \int_0^s e^{-\boldsymbol{\Theta}(t-u)}\boldsymbol{\Sigma}^{\frac{1}{2}} d\mathbf{W}_u ds$ is independent of $\mathbf{y}_0$, we have that

$$Cov\left[\bar{\mathbf{y}}_{0,t}, \mathbf{y}_0\right] = Cov\left[\frac{1}{t}\boldsymbol{\Theta}^{-1}\left(e^{-\boldsymbol{\Theta}t} - \mathbf{I}\right)\left(\boldsymbol{\mu} - \mathbf{y}_0\right), \mathbf{y}_0\right] \tag{A83}$$

$$= -\frac{1}{t}\boldsymbol{\Theta}^{-1}\left(e^{-\boldsymbol{\Theta}t} - \mathbf{I}\right) Cov\left[\mathbf{y}_0, \mathbf{y}_0\right] \tag{A84}$$

$$= -\frac{1}{t}\boldsymbol{\Theta}^{-1}\left(e^{-\boldsymbol{\Theta}t} - \mathbf{I}\right) Var\left[\mathbf{y}_0\right]. \tag{A85}$$

Covariance between $\mathbf{y}_0$ *and* $\mathbf{y}_t$ Since

$$\mathbf{y}_t = e^{-\boldsymbol{\Theta}t}\mathbf{y}_0 + \left(\mathbf{I} - e^{-\boldsymbol{\Theta}t}\right)\boldsymbol{\mu} + \int_0^t e^{-\boldsymbol{\Theta}(t-s)}\boldsymbol{\Sigma}^{\frac{1}{2}} d\mathbf{W}_s \tag{A86}$$

and $\int_0^t e^{-\boldsymbol{\Theta}(t-s)}\boldsymbol{\Sigma}^{\frac{1}{2}}d\mathbf{W}_s$ is independent of $\mathbf{y}_0$, we have that

$$Cov\left[\bar{\mathbf{y}}_{0,t},\mathbf{y}_0\right] = Cov\left[e^{-\boldsymbol{\Theta}t},\mathbf{y}_0\right] \tag{A87}$$

$$= e^{-\boldsymbol{\Theta}t}Cov\left[\mathbf{y}_0,\mathbf{y}_0\right] \tag{A88}$$

$$= e^{-\boldsymbol{\Theta}t}Var\left[\mathbf{y}_0\right]. \tag{A89}$$

Covariance between $\bar{\mathbf{y}}_{0,t}$ *and* $\mathbf{y}_t$ Since

$$\bar{\mathbf{y}}_{0,t} = \frac{1}{t}\boldsymbol{\Theta}^{-1}\left(e^{-\boldsymbol{\Theta}t}-\mathbf{I}\right)\left(\boldsymbol{\mu}-\mathbf{y}_0\right)+\boldsymbol{\mu}+\frac{1}{t}\int_0^t\int_0^s e^{-\boldsymbol{\Theta}(s-u)}\boldsymbol{\Sigma}^{\frac{1}{2}}d\mathbf{W}_u ds \tag{A90}$$

and

$$\mathbf{y}_t = e^{-\boldsymbol{\Theta}t}\mathbf{y}_0+\left(\mathbf{I}-e^{-\boldsymbol{\Theta}t}\right)\boldsymbol{\mu}+\int_0^t e^{-\boldsymbol{\Theta}(t-s)}\boldsymbol{\Sigma}^{\frac{1}{2}}d\mathbf{W}_s \tag{A91}$$

we have

$$Cov\left[\bar{\mathbf{y}}_{0,t},\mathbf{y}_t\right] = Cov\left[\frac{1}{t}\boldsymbol{\Theta}^{-1}\left(e^{-\boldsymbol{\Theta}t}-\mathbf{I}\right)\left(\boldsymbol{\mu}-\mathbf{y}_0\right),\mathbf{y}_t\right]+Cov\left[\frac{1}{t}\int_0^t\int_0^s e^{-\boldsymbol{\Theta}(s-u)}\boldsymbol{\Sigma}^{\frac{1}{2}}d\mathbf{W}_u ds,\mathbf{y}_t\right] \tag{A92}$$

$$\begin{aligned}= &-\frac{1}{t}\boldsymbol{\Theta}^{-1}\left(e^{-\boldsymbol{\Theta}t}-\mathbf{I}\right)Cov\left[\mathbf{y}_0,e^{-\boldsymbol{\Theta}t}\mathbf{y}_0+\left(\mathbf{I}-e^{-\boldsymbol{\Theta}t}\right)\boldsymbol{\mu}+\int_0^t e^{-\boldsymbol{\Theta}(t-s)}\boldsymbol{\Sigma}^{\frac{1}{2}}d\mathbf{W}_s\right]\\ &+Cov\left[\frac{1}{t}\int_0^t\int_0^s e^{-\boldsymbol{\Theta}(s-u)}\boldsymbol{\Sigma}^{\frac{1}{2}}d\mathbf{W}_u ds,\mathbf{y}_t\right]\end{aligned} \tag{A93}$$

$$\begin{aligned}= &-\frac{1}{t}\boldsymbol{\Theta}^{-1}\left(e^{-\boldsymbol{\Theta}t}-\mathbf{I}\right)Cov\left[\mathbf{y}_0,e^{-\boldsymbol{\Theta}t}\mathbf{y}_0\right]\\ &-\frac{1}{t}\boldsymbol{\Theta}^{-1}\left(e^{-\boldsymbol{\Theta}t}-\mathbf{I}\right)Cov\left[\mathbf{y}_0,\left(\mathbf{I}-e^{-\boldsymbol{\Theta}t}\right)\boldsymbol{\mu}+\int_0^t e^{-\boldsymbol{\Theta}(t-s)}\boldsymbol{\Sigma}^{\frac{1}{2}}d\mathbf{W}_s\right]\\ &+Cov\left[\frac{1}{t}\int_0^t\int_0^s e^{-\boldsymbol{\Theta}(s-u)}\boldsymbol{\Sigma}^{\frac{1}{2}}d\mathbf{W}_u ds,\mathbf{y}_t\right].\end{aligned} \tag{A94}$$

Since $\left(\mathbf{I} - e^{-\mathbf{\Theta} t}\right) \boldsymbol{\mu}$ is a constant and $\int_0^t e^{-\mathbf{\Theta}(t-s)} \mathbf{\Sigma}^{\frac{1}{2}} d\mathbf{W}_s$ is independent of $\mathbf{y}_0$, we have

$$
\begin{aligned}
Cov\left[\bar{\mathbf{y}}_{0,t}, \mathbf{y}_t\right] &= -\frac{1}{t}\mathbf{\Theta}^{-1}\left(e^{-\mathbf{\Theta} t} - \mathbf{I}\right) Cov\left[\mathbf{y}_0, e^{-\mathbf{\Theta} t}\mathbf{y}_0\right] \\
&\quad + Cov\left[\frac{1}{t}\int_0^t \int_0^s e^{-\mathbf{\Theta}(s-u)} \mathbf{\Sigma}^{\frac{1}{2}} d\mathbf{W}_u ds, \mathbf{y}_t\right] \qquad \text{(A95)} \\
&= -\frac{1}{t}\mathbf{\Theta}^{-1}\left(e^{-\mathbf{\Theta} t} - \mathbf{I}\right) Cov\left[\mathbf{y}_0, \mathbf{y}_0\right] e^{-\mathbf{\Theta}^{\mathrm{T}} t} \\
&\quad + Cov\left[\frac{1}{t}\int_0^t \int_0^s e^{-\mathbf{\Theta}(s-u)} \mathbf{\Sigma}^{\frac{1}{2}} d\mathbf{W}_u ds, e^{-\mathbf{\Theta} t}\mathbf{y}_0 + \left(\mathbf{I} - e^{-\mathbf{\Theta} t}\right)\boldsymbol{\mu} + \int_0^t e^{-\mathbf{\Theta}(t-s)} \mathbf{\Sigma}^{\frac{1}{2}} d\mathbf{W}_s\right] \qquad \text{(A96)} \\
&= -\frac{1}{t}\mathbf{\Theta}^{-1}\left(e^{-\mathbf{\Theta} t} - \mathbf{I}\right) Var\left[\mathbf{y}_0\right] e^{-\mathbf{\Theta}^{\mathrm{T}} t} \\
&\quad + Cov\left[\frac{1}{t}\int_0^t \int_0^s e^{-\mathbf{\Theta}(s-u)} \mathbf{\Sigma}^{\frac{1}{2}} d\mathbf{W}_u ds, e^{-\mathbf{\Theta} t}\mathbf{y}_0\right] \qquad \text{(A97)} \\
&\quad + Cov\left[\frac{1}{t}\int_0^t \int_0^s e^{-\mathbf{\Theta}(s-u)} \mathbf{\Sigma}^{\frac{1}{2}} d\mathbf{W}_u ds, \int_0^t e^{-\mathbf{\Theta}(t-s)} \mathbf{\Sigma}^{\frac{1}{2}} d\mathbf{W}_s\right].
\end{aligned}
$$

Similarly noting that $\frac{1}{t}\int_0^t \int_0^s e^{-\mathbf{\Theta}(t-u)} \mathbf{\Sigma}^{\frac{1}{2}} d\mathbf{W}_u ds$ is independent of $\mathbf{y}_0$ yields

$$
Cov\left[\bar{\mathbf{y}}_{0,t}, \mathbf{y}_t\right] = -\frac{1}{t}\mathbf{\Theta}^{-1}\left(e^{-\mathbf{\Theta} t} - \mathbf{I}\right) Var\left[\mathbf{y}_0\right] e^{-\mathbf{\Theta}^{\mathrm{T}} t} + Cov\left[\frac{1}{t}\int_0^t \int_0^s e^{-\mathbf{\Theta}(s-u)} \mathbf{\Sigma}^{\frac{1}{2}} d\mathbf{W}_u ds, \int_0^t e^{-\mathbf{\Theta}(t-s)} \mathbf{\Sigma}^{\frac{1}{2}} d\mathbf{W}_s\right]. \qquad \text{(A98)}
$$

Since $E\left[\frac{1}{t}\int_0^t \int_0^s e^{-\mathbf{\Theta}(s-u)} \mathbf{\Sigma}^{\frac{1}{2}} d\mathbf{W}_u ds\right] = 0$ and $E\left[\int_0^t e^{-\mathbf{\Theta}(t-s)} \mathbf{\Sigma}^{\frac{1}{2}} d\mathbf{W}_s\right] = 0$, we find that

$$
\begin{aligned}
&Cov\left[\frac{1}{t}\int_0^t \int_0^s e^{-\mathbf{\Theta}(s-u)} \mathbf{\Sigma}^{\frac{1}{2}} d\mathbf{W}_u ds, \int_0^t e^{-\mathbf{\Theta}(t-s)} \mathbf{\Sigma}^{\frac{1}{2}} d\mathbf{W}_s\right] \qquad \text{(A99)} \\
&\quad = E\left[\frac{1}{t}\int_0^t \int_0^s e^{-\mathbf{\Theta}(s-u)} \mathbf{\Sigma}^{\frac{1}{2}} d\mathbf{W}_u ds \int_0^t \left\{d\mathbf{W}_v^{\mathrm{T}} \mathbf{\Sigma}^{\frac{1}{2}} e^{-\mathbf{\Theta}^{\mathrm{T}}(t-v)}\right\}\right] \qquad \text{(A100)} \\
&\quad = \frac{1}{t} E\left[\int_0^t \int_0^t \int_0^s \left\{e^{-\mathbf{\Theta}(s-u)} \mathbf{\Sigma}^{\frac{1}{2}} d\mathbf{W}_u d\mathbf{W}_v^{\mathrm{T}} \left(\mathbf{\Sigma}^{\frac{1}{2}}\right)^{\mathrm{T}} e^{-\mathbf{\Theta}^{\mathrm{T}}(t-v)}\right\} ds\right]. \qquad \text{(A101)}
\end{aligned}
$$

As before, $E\left[\int_0^t e^{-\mathbf{\Theta}(t-s)} \mathbf{\Sigma}^{\frac{1}{2}} d\mathbf{W}_s\right] = 0$, $d\mathbf{W}_u d\mathbf{W}_u^{\mathrm{T}} = I du$, and $d\mathbf{W}_u d\mathbf{W}_v^{\mathrm{T}} = 0$ for $u \neq v$. Thus,

$$
\begin{aligned}
&Cov\left[\frac{1}{t}\int_0^t \int_0^s e^{-\mathbf{\Theta}(t-u)} \mathbf{\Sigma}^{\frac{1}{2}} d\mathbf{W}_u ds, \int_0^t e^{-\mathbf{\Theta}(t-s)} \mathbf{\Sigma}^{\frac{1}{2}} d\mathbf{W}_s\right] \qquad \text{(A102)} \\
&\quad = \frac{1}{t} E\left[\int_0^t \int_0^s \left\{e^{-\mathbf{\Theta}(s-u)} \mathbf{\Sigma}^{\frac{1}{2}} du \left(\mathbf{\Sigma}^{\frac{1}{2}}\right)^{\mathrm{T}} e^{-\mathbf{\Theta}^{\mathrm{T}}(t-u)}\right\} ds\right] \qquad \text{(A103)} \\
&\quad = \frac{1}{t}\int_0^t \int_0^s \left\{e^{-\mathbf{\Theta}(s-u)} \mathbf{\Sigma} e^{-\mathbf{\Theta}^{\mathrm{T}}(t-u)}\right\} du\, ds. \qquad \text{(A104)}
\end{aligned}
$$

Noting $(\boldsymbol{\Theta} \oplus \boldsymbol{\Theta})^{-1}$ and $\left(\mathbf{I} \otimes e^{-\boldsymbol{\Theta}(t-s)}\right)$ commute and vectorizing yields

$$\text{vec}\left(Cov\left[\frac{1}{t}\int_0^t \int_0^s e^{-\boldsymbol{\Theta}(s-u)}\boldsymbol{\Sigma}^{\frac{1}{2}} d\mathbf{W}_u ds, \int_0^t e^{-\boldsymbol{\Theta}(t-s)}\boldsymbol{\Sigma}^{\frac{1}{2}} d\mathbf{W}_s\right]\right) \tag{A105}$$

$$= \frac{1}{t}\text{vec}\left(\int_0^t \int_0^s \left\{e^{-\boldsymbol{\Theta}(s-u)}\boldsymbol{\Sigma} e^{-\boldsymbol{\Theta}^{\top}(t-u)}\right\} du\, ds\right) \tag{A106}$$

$$= \frac{1}{t}\int_0^t \int_0^s e^{-\boldsymbol{\Theta}(s-u)} \otimes e^{-\boldsymbol{\Theta}(t-u)} du\, ds\, \text{vec}\left(\boldsymbol{\Sigma}\right) \tag{A107}$$

$$= \frac{1}{t}\int_0^t \left(\mathbf{I} \otimes e^{-\boldsymbol{\Theta}(t-s)}\right) \int_0^s \left(e^{-\boldsymbol{\Theta}(s-u)} \otimes e^{-\boldsymbol{\Theta}(s-u)}\right) du\, ds\, \text{vec}\left(\boldsymbol{\Sigma}\right) \tag{A108}$$

$$= \frac{1}{t}\int_0^t \left(\mathbf{I} \otimes e^{-\boldsymbol{\Theta}(t-s)}\right) \int_0^s e^{-(\boldsymbol{\Theta}\oplus\boldsymbol{\Theta})(s-u)} du\, ds\, \text{vec}\left(\boldsymbol{\Sigma}\right) \tag{A109}$$

$$= \frac{1}{t}\int_0^t \left(\mathbf{I} \otimes e^{-\boldsymbol{\Theta}(t-s)}\right) (\boldsymbol{\Theta} \oplus \boldsymbol{\Theta})^{-1} \left(\mathbf{I} - e^{-(\boldsymbol{\Theta}\oplus\boldsymbol{\Theta})s}\right) ds\, \text{vec}\left(\boldsymbol{\Sigma}\right) \tag{A110}$$

$$= \frac{1}{t}(\boldsymbol{\Theta} \oplus \boldsymbol{\Theta})^{-1} \int_0^t \left(\mathbf{I} \otimes e^{-\boldsymbol{\Theta}(t-s)}\right) \left(\mathbf{I} - e^{-(\boldsymbol{\Theta}\oplus\boldsymbol{\Theta})s}\right) ds\, \text{vec}\left(\boldsymbol{\Sigma}\right) \tag{A111}$$

$$\begin{aligned} &= \frac{1}{t}(\boldsymbol{\Theta} \oplus \boldsymbol{\Theta})^{-1} \int_0^t \left(\mathbf{I} \otimes e^{-\boldsymbol{\Theta}(t-s)}\right) ds\, \text{vec}\left(\boldsymbol{\Sigma}\right) \\ &\quad + \frac{1}{t}(\boldsymbol{\Theta} \oplus \boldsymbol{\Theta})^{-1} \int_0^t \left(\mathbf{I} \otimes e^{-\boldsymbol{\Theta}(t-s)}\right) \left(e^{-\boldsymbol{\Theta}s} \otimes e^{-\boldsymbol{\Theta}s}\right) ds\, \text{vec}\left(\boldsymbol{\Sigma}\right) \end{aligned} \tag{A112}$$

$$\begin{aligned} &= \frac{1}{t}(\boldsymbol{\Theta} \oplus \boldsymbol{\Theta})^{-1} \left(\mathbf{I} \otimes \int_0^t e^{-\boldsymbol{\Theta}(t-s)} ds\right) \text{vec}\left(\boldsymbol{\Sigma}\right) \\ &\quad - \frac{1}{t}(\boldsymbol{\Theta} \oplus \boldsymbol{\Theta})^{-1} \left(\int_0^t e^{-\boldsymbol{\Theta}s} ds \otimes e^{-\boldsymbol{\Theta}t}\right) \text{vec}\left(\boldsymbol{\Sigma}\right) \end{aligned} \tag{A113}$$

$$\begin{aligned} &= \frac{1}{t}(\boldsymbol{\Theta} \oplus \boldsymbol{\Theta})^{-1} \left(\mathbf{I} \otimes \boldsymbol{\Theta}^{-1}\left(\mathbf{I} - e^{-\boldsymbol{\Theta}t}\right)\right) \text{vec}\left(\boldsymbol{\Sigma}\right) \\ &\quad \frac{1}{t}(\boldsymbol{\Theta} \oplus \boldsymbol{\Theta})^{-1} \left(\boldsymbol{\Theta}^{-1}\left(e^{-\boldsymbol{\Theta}t} - \mathbf{I}\right) \otimes e^{-\boldsymbol{\Theta}t}\right) \text{vec}\left(\boldsymbol{\Sigma}\right) \end{aligned} \tag{A114}$$

$$\begin{aligned} &= \frac{1}{t}(\boldsymbol{\Theta} \oplus \boldsymbol{\Theta})^{-1} \left(\mathbf{I} \otimes \boldsymbol{\Theta}^{-1}\left(\mathbf{I} - e^{-\boldsymbol{\Theta}t}\right)\right) \text{vec}\left(\boldsymbol{\Sigma}\right) \\ &\quad - \frac{1}{t}(\boldsymbol{\Theta} \oplus \boldsymbol{\Theta})^{-1} \left(\boldsymbol{\Theta}^{-1}\left(\mathbf{I} - e^{-\boldsymbol{\Theta}t}\right) \otimes e^{-\boldsymbol{\Theta}t}\right) \text{vec}\left(\boldsymbol{\Sigma}\right) \end{aligned} \tag{A115}$$

$$= \frac{1}{t}(\boldsymbol{\Theta} \oplus \boldsymbol{\Theta})^{-1} \left(\left(\mathbf{I} \otimes \boldsymbol{\Theta}^{-1}\left(\mathbf{I} - e^{-\boldsymbol{\Theta}t}\right)\right) + \left(\boldsymbol{\Theta}^{-1}\left(\mathbf{I} - e^{-\boldsymbol{\Theta}t}\right) \otimes e^{-\boldsymbol{\Theta}t}\right)\right) \text{vec}\left(\boldsymbol{\Sigma}\right) \tag{A116}$$

Thus, substituting Equation A116 into Equation A98, we find that

$$\begin{aligned} \text{vec}\left(Cov\left[\bar{\mathbf{y}}_{0,t}, \mathbf{y}_t\right]\right) &= \text{vec}\left(-\frac{1}{t}\boldsymbol{\Theta}^{-1}\left(e^{-\boldsymbol{\Theta}t} - \mathbf{I}\right) Var\left[\mathbf{y}_0\right] e^{-\boldsymbol{\Theta}^{\top}t}\right) \\ &\quad + \frac{1}{t}(\boldsymbol{\Theta} \oplus \boldsymbol{\Theta})^{-1} \left(\mathbf{I} \otimes \boldsymbol{\Theta}^{-1}\left(\mathbf{I} - e^{-\boldsymbol{\Theta}t}\right)\boldsymbol{\Theta}^{-1}\left(\mathbf{I} - e^{-\boldsymbol{\Theta}t}\right) \otimes e^{-\boldsymbol{\Theta}t}\right) \text{vec}\left(\boldsymbol{\Sigma}\right) \end{aligned} \tag{A117}$$

$$\begin{aligned} &= -\frac{1}{t}\left(\mathbf{I} \otimes \boldsymbol{\Theta}^{-1}\right)\left(e^{-(\boldsymbol{\Theta}\oplus\boldsymbol{\Theta})t} - e^{-\boldsymbol{\Theta}t} \otimes \mathbf{I}\right) \text{vec}\left(Var\left[\mathbf{y}_0\right]\right) \\ &\quad + \frac{1}{t}(\boldsymbol{\Theta} \oplus \boldsymbol{\Theta})^{-1} \left(\mathbf{I} \otimes \boldsymbol{\Theta}^{-1}\left(\mathbf{I} - e^{-\boldsymbol{\Theta}t}\right)\boldsymbol{\Theta}^{-1}\left(\mathbf{I} - e^{-\boldsymbol{\Theta}t}\right) \otimes e^{-\boldsymbol{\Theta}t}\right) \text{vec}\left(\boldsymbol{\Sigma}\right). \end{aligned} \tag{A118}$$

B.4. *Stationary distribution*

From Equation A14, we have

$$\mathbf{y}_t = e^{-\boldsymbol{\Theta} t}\mathbf{y}_0 + \left(\mathbf{I} - e^{-\boldsymbol{\Theta} t}\right)\boldsymbol{\mu} + \int_0^t e^{-\boldsymbol{\Theta}(t-s)}\boldsymbol{\Sigma}^{\frac{1}{2}} d\mathbf{W}_s. \tag{A119}$$

Since $\int_0^t e^{-\boldsymbol{\Theta}(t-s)}\boldsymbol{\Sigma}^{\frac{1}{2}} d\mathbf{W}_s$ is a linear operator on a Wiener process, it must follow a multivariate normal distribution. As before, note that $E\left[\int_0^t e^{-\boldsymbol{\Theta}(t-s)}\boldsymbol{\Sigma}^{\frac{1}{2}} d\mathbf{W}_s\right] = 0$. Equation A24 established that

$$\operatorname{vec}\left(Var\left[\int_0^t e^{-\boldsymbol{\Theta}(t-s)}\boldsymbol{\Sigma}^{\frac{1}{2}} d\mathbf{W}_s\right]\right) = (\boldsymbol{\Theta}\oplus\boldsymbol{\Theta})^{-1}\left(\mathbf{I} - e^{-(\boldsymbol{\Theta}\oplus\boldsymbol{\Theta})t}\right)\operatorname{vec}(\boldsymbol{\Sigma}) \tag{A120}$$

$$\lim_{t\to\infty}\operatorname{vec}\left(Var\left[\int_0^t e^{-\boldsymbol{\Theta}(t-s)}\boldsymbol{\Sigma}^{\frac{1}{2}} d\mathbf{W}_s\right]\right) = (\boldsymbol{\Theta}\oplus\boldsymbol{\Theta})^{-1}\operatorname{vec}(\boldsymbol{\Sigma}). \tag{A121}$$

Define $\mathbf{P}$ by $\operatorname{vec}(\mathbf{P}) = (\boldsymbol{\Theta}\oplus\boldsymbol{\Theta})^{-1}\operatorname{vec}(\boldsymbol{\Sigma})$. Thus,

$$\int_0^t e^{-\boldsymbol{\Theta}(t-s)}\boldsymbol{\Sigma}^{\frac{1}{2}} d\mathbf{W}_s \xrightarrow{d} \mathcal{N}(0,\mathbf{P}). \tag{A122}$$

Note that $\lim_{n\to\infty}\left(\mathbf{I} - e^{-\boldsymbol{\Theta} t}\right)\boldsymbol{\mu} = \boldsymbol{\mu}$. Finally, since the real part of the eigenvalues of $\boldsymbol{\Theta}$ are positive by assumption, $e^{-\boldsymbol{\Theta} t}\mathbf{y}_0$ converges pointwise to 0 as $t\to\infty$. It follows from Slutsky's theorem that

$$\mathbf{y}_t \xrightarrow{d} \mathcal{N}(\boldsymbol{\mu},\mathbf{P}) \tag{A123}$$

as $t\to\infty$.

C. HIGHER-ORDER MODELS

Although the latent process underlying the MONOCAR model has been assumed to be a first-order multivariate continuous-time, or MCAR(1) process, the MONOCAR model can be extended to use a higher-order MCAR(p) process, the continuous-time analogs of a VAR(p) process. Higher-order continuous-time processes can be thought of as including velocity or acceleration terms, thus allowing us to capture processes in which, for example, a trending series tends to continue trending. For example, a latent process which is thought of as having "momentum" might be well described by a MONOCAR(2) model.

A MONOCAR(p) model can be described by generalizing the equations defining a MONOCAR. The stochastic differential equation defining the multivariate Ornstein-Uhlenbeck process which describes the latent process in a MONOCAR model can be written as

$$D\mathbf{y}_t + \boldsymbol{\Theta}\mathbf{y}_t = \boldsymbol{\Theta}\boldsymbol{\mu} + \Sigma^{\frac{1}{2}} D\mathbf{W}_t, \qquad D \equiv \frac{d}{dt}, \tag{A124}$$

where $\mathbf{W}_t$ is a multivariate Wiener process. Following Schlemm and Stelzer (2012), we can extend this to a p^{th}-order process defined by the stochastic differential equation

$$D^p\mathbf{y}_t + \tilde{\boldsymbol{\Theta}}_1 D^{p-1}\mathbf{y}_t + \cdots + \tilde{\boldsymbol{\Theta}}_{p-1} D\mathbf{y}_t + \tilde{\boldsymbol{\Theta}}_p \mathbf{y}_t = \tilde{\boldsymbol{\Theta}}_p \mu + \Sigma^{\frac{1}{2}} D\mathbf{W}_t, \tag{A125}$$

with all other components of the MONOCAR model unchanged.

Much as a VAR(p) model can be represented as a VAR(1) model, a higher-order MONOCAR model can also be represented as special case of the first-order MONOCAR model described above. A MONOCAR(p) model with k latent processes is equivalent to a first-order,

pk-dimensional model with

$$\boldsymbol{\Theta} = \begin{pmatrix} 0 & \mathbf{I}_k & 0 & \cdots & 0 \\ 0 & 0 & \mathbf{I}_k & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & 0 & \mathbf{I}_k \\ \tilde{\boldsymbol{\Theta}}_p & \tilde{\boldsymbol{\Theta}}_{p-1} & \tilde{\boldsymbol{\Theta}}_{p-2} & \cdots & \tilde{\boldsymbol{\Theta}}_1 \end{pmatrix} \tag{A126}$$

$$\boldsymbol{\Sigma} = \begin{pmatrix} 0 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & 0 \\ 0 & \cdots & 0 & 0 \\ 0 & \cdots & 0 & \tilde{\boldsymbol{\Sigma}} \end{pmatrix} \tag{A127}$$

where $\mathbf{I}_k$ is the $k \times k$ identity matrix, $\boldsymbol{\Theta}$ and $\boldsymbol{\Sigma}$ are $pk \times pk$ matrices, $\tilde{\boldsymbol{\Theta}}_1, \ldots, \tilde{\boldsymbol{\Theta}}_p$ are $k \times k$ matrices of parameters capturing the relationship between the latent series, and $\tilde{\boldsymbol{\Sigma}}$ is the $k \times k$ matrix of unknown variance parameters.[6] Using this representation, a MONOCAR(p) model can be estimated within the current software.[7]

In considering higher-order MONOCAR models, the selection of an appropriate order can be done by minimizing the Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC). Other criteria, such as cross-validated error, could also be used. A hypothesis test of a MONOCAR(p) model against a MONOCAR(p') model can be performed using a likelihood ratio test. In some cases, there may be theoretical reasons for considering a MONOCAR model of a particular order.

Despite the availability of higher-order MONOCAR models, low-order—often MONOCAR(1)—models are often sufficient. The inclusion of measurement error can allow a low-order MONOCAR to capture effects that would necessitate a higher-order VAR model. Low-order

[6]Observations remain a function of the same components of the latent process as in the original MONOCAR(p) representation. Thus, observations are only a function of the first k dimensions of this pk-dimensional process.

[7]Further extensions to MCARMA(p, q) processes are also possible in this same framework. See, e.g., Schlemm and Stelzer (2012) for details on similar representations of MCARMA(p, q) processes.

MONOCAR models also offer greater interpretability of coefficients.

D. HOUSE EFFECTS

Polling data is often drawn from opinion polls conducted by multiple polling organizations or from answers to questions with slight variations in wording. These differences can impact the responses and ignoring these differences can be dangerous. On the other hand, data collected by a single organization, using a precise question wording, and differing in no other relevant respects may be very limited and these difference may be relatively small.

A common solution is to introduce a term for the bias of each polling organization, termed a "house effect" (see Jackman 2005). Thus, the polls conducted by one organization are assumed to differ from the polls conducted by another organization by a constant. This same approach can be used to capture "question wording effects" due to small differences in question wording between polls.

This approach avoids the need to treat the polls from different organizations as entirely different series while still allowing the model to account for differences between organizations. In contrast, treating these polls as different series would necessitate adding many more parameters to the model and complicating inference. Moreover, house effects are often sufficient to account for the differences between polls, particularly when the differences are relatively small. This approach can be extended to permit the variance transformation discussed in the previous section to vary by polling organization, allowing for the possibility that some organizations produce results with lower variance than other organizations.

E. VARIANCE-STABILIZING TRANSFORMATIONS

Variance-stabilizing transformations have a long history in statistical practice, particularly before modern computational resources made non-Gaussian models practical. These transfor-

mations result in distributions with approximately constant variance and, asymptotically, Gaussian distributions with exactly constant variance. Such transformations can be derived for many standard distributions.

Two common case that would require variance stabilizing transformations are the binomial and Poisson distributions. The canonical variance-stabilizing transformations are the angular transformation $z' = \arcsin\left(\sqrt{\frac{z}{N}}\right)$ for the binomial and the square-root function, $z' = \sqrt{z}$ for the Poisson. Several improved versions of the angular transformation of binomial data have been proposed, such as Anscombe's (1948) proposal of the form $\arcsin\sqrt{(z+c)/(N+2c)}$. However, such transformations are less appropriate because of they depend on n. Chanter (1975) addresses precisely this issue and proposes an improved transformation that does not vary with n.

Chanter (1975) addresses precisely this issue and proposes an improved transformation that does not vary with n. This work suggests the transformation

$$z' = \arcsin\left(\left(1 - \tfrac{1}{5000}\right)\left(2\frac{z}{N} - 1\right)\right) \tag{A128}$$

for the binomial distribution and

$$z' = \sqrt{z + \frac{3}{8}} \tag{A129}$$

for the Poisson distribution.[8] Their respective approximate variances are $\frac{1}{N}$ and $\frac{1}{4}$. These transformations are proposed in the paper instead of the canonical transformations because they perform slightly better in most settings.

[8]The representation of the transformation for the binomial distribution is slightly different than the representation in Chanter (1975) but is mathematically equivalent. This equivalence follows from applying the trigonometric identity $2\arcsin\sqrt{\frac{z}{N}} = \arcsin(2\frac{z}{N} - 1) + \pi/2$.

E.1. *Alternatives to variance-stabilizing transformations*

Most other solution to the complications addressed by variance-stabilizing transformations involve modifying the model. For example, in the distributional assumptions of our observation process, the variance might be allowed to depend on the mean or a binomial (or other) distribution might be used in place of a Gaussian distribution. Similarly, the bounds on the data might be addressed by transforming the latent variable or replacing the Ornstein-Uhlenbeck process with a bounded alternative, such as a multivariate Jacobi process.

One simple alternative would be to drop the use of variance-stabilizing transformations and instead use an extended or unscented Kalman filter for approximate likelihood computation. This would have the advantage of allowing the parameters to be interpreted in terms of natural units. However, in addition to eliminating the possibility of exact likelihood computation, this approach would also exacerbate problems with bounded variables.

The use of a Markov-chain Monte Carlo (MCMC) methods in a Bayesian framework would also be possible, but is not quite as straightforward as in many other applications. Because the latent process is infinite-dimensional, it is not possible to treat the entire process as an auxiliary variable, as can be done in many other cases. We may also prefer to remain in a frequentist framework. Nonetheless, MCMC methods can be applied to the MONOCAR model and may offer certain advantages, such as allowing computation without the need for a Gaussian approximation of the sampling distribution.

Almost all alternatives depend on either an approximation to the likelihood or the use of Monte Carlo methods. This is not always practical and the use of an approximation to the likelihood is not necessarily superior to an approximation to the distributional form.

F. STRUCTURAL MONOCAR

In some circumstances, a more general parameterization of the MONOCAR model may be necessary to provide a behavioral interpretation of the parameters, analogous to the structural form of a VAR. However before considering a structural form of a MONOCAR model beyond the parameterization that has been used so far, which might be termed the reduced form, it is worth noting that the structural form of a MONOCAR model already captures one form of contemporaneous effects between variables. As a result, it can be appropriate to assume that the structural form of a MONOCAR model is the reduced form even when this would not be true for VAR models.

An important reason for using a structural VAR is to identify contemporaneous relationships between the variables. Assuming that the structural form of a VAR model is the reduced form, implicitly assumes that no contemporaneous relationships exist. MONOCAR, however, already describes contemporaneous relationships between the variables. In a MONOCAR model, the value of each series has a contemporaneous effect on the rate of change of each other series. Thus, while assuming that the structural form of a VAR model is the reduced form implicitly assumes that there are no contemporaneous relationships between the variables, we make no such assumption by using a MONOCAR model.

Consider a MONOCAR model with independent error processes observed at regular monthly intervals. Because the underlying latent process acts continuously with time, the processes continue to affect each other between observations. As a result, the data would include a contemporaneous relationship from a discrete-time perspective. If we were analyzing this data with vector autoregression, a structural VAR approach would be necessary to identify these contemporaneous effects. However, a standard MONOCAR model would already capture these relationships.

We can also find a more formal equivalence between the parameters of a structural VAR and a special case of the MONOCAR. Suppose the data are generated by a MONOCAR

model with instantaneous observations at regular intervals and no measurement error. For simplicity, assume that the long-run mean, μ, is zero. Thus, the latent process is described by

$$dy_t = \Theta\left(\mu - y_t\right) dt + \Sigma^{\frac{1}{2}} dW_t. \tag{A130}$$

However, as Bergstrom (1988) points out, this process is also is well approximated by the structural VAR model given by

$$\left(\mathbf{I} + \frac{1}{2}\Theta\right) y_t = \Theta\mu + \frac{1}{2}\left(\mathbf{I} - \frac{1}{2}\Theta\right)\Theta y_{t-1} + \varepsilon_t, \tag{A131}$$

which can be written as

$$\Delta y_t = \Theta\left(\mu - \frac{y_t + y_{t-1}}{2}\right) + \varepsilon_t. \tag{A132}$$

This parameterization is virtually identical to the parameterization of the MONOCAR model, whereas the reduced-form parameterization is not.

However, if one wishes to identify both instantaneous and gradual effects in a MONOCAR model, a structural form of the model would be useful. Much as a structural VAR provides a more general parameterization, a similar general parameterization can be provided for the MONOCAR model. The structural MONOCAR can be constructed by replacing equation 3 by

$$\mathbf{H} d\mathbf{y}_t = \boldsymbol{\Theta}\left(\boldsymbol{\mu} - \mathbf{y}_t\right) dt + \boldsymbol{\Sigma}^{\frac{1}{2}} d\mathbf{W}_t, \tag{A133}$$

with the corresponding reduced form given by

$$d\mathbf{y}_t = \mathbf{H}^{-}1\boldsymbol{\Theta}\left(\boldsymbol{\mu} - \mathbf{y}_t\right) dt + \mathbf{H}^{-}1\boldsymbol{\Sigma}^{\frac{1}{2}} d\mathbf{W}_t. \tag{A134}$$

By treating the reduced form of a MONOCAR model as the structural model, we are implicitly assuming that $\mathbf{H} = \mathbf{I}$. As with structural VAR, the structural MONOCAR model is unidentified without restrictions on the parameters in equation A133.

The structural form of a MONOCAR differs from the reduced form by by capturing non-gradual contemporaneous effects. Unlike the reduced form of a VAR model, the reduced form of a MONOCAR model already treats each series as potentially having a contemporaneous effect on the other series. However, these effects are gradual; while there is no lag before a shock to one series begins to affect another series, the size of this effect grows with time. A structural MONOCAR can also capture contemporaneous effects in which the full effect of a shock to one series appears immediately in another series.

G. MORE DETAILS ON MODELS FROM SECTION

G.1. *State-space models*

State-space models provide another possible approach to handling measurement error and missing values in time-series data. State-space models are a class of models that explicitly model observations as a function of latent processes. Various types of state-space models have gained wide usage in engineering, biology, physics, and economics. However, state-space models have garnered relatively little usage in political science.[9]

State-space models consist of a transition equation and an observation equation.[10] The transition equation captures the evolution of the state variable over time but is not directly observed. The observation equation links the observations at time t to the state at time t. When the latent variables are discrete, the models are more commonly known as hidden Markov models. The term state-space model is sometimes used to refer only to cases where the state space—that is, the set of possible values of the state variable—is continuous. Note that this distinction is separate from whether time is continuous; time may still be discrete

[9]Important exceptions include Beck (1989); Green, Gerber, and Boef (1999); McAvoy (1998); Kellstedt, McAvoy, and Stimson (1993); Brandt et al. (2000); Brandt and Williams (2001); Martin and Quinn (2002); Bond, Fleisher, and Wood (2003); Jackman (2005); Pickup and Johnston (2007); De Boef and Keele (2008); Armstrong (2009).

[10]See Commandeur and Koopman (2007) for a more thorough introduction to MARSS and other state-space models and Durbin and Koopman (2012) for a more advanced treatment.

in state-space models where the state space is continuous.

State-space models allow us to model a vector autoregressive process that is not directly observed. Rather, each observation can be modeled as a noisy measure of the latent process or a linear function of the latent process at some point in time.[11] The latent process is generally assumed to be a first-order vector autoregressive process, although such a model can be used to capture a higher-order vector autoregressive process by expanding the state space. Thus, the model becomes

$$\mathbf{y}_t = \boldsymbol{\alpha} + \mathbf{B}\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t \tag{A135}$$

$$\mathbf{z}_t = \boldsymbol{\eta} + \mathbf{Z}\mathbf{y}_t + \boldsymbol{\omega}_t \tag{A136}$$

with

$$\boldsymbol{\varepsilon}_t \overset{iid}{\sim} \mathcal{N}(0, \mathbf{T}) \tag{A137}$$

$$\boldsymbol{\omega}_t \overset{iid}{\sim} \mathcal{N}(0, \mathbf{P}) \tag{A138}$$

where $\mathbf{z}_t$ is an observation at time t, and $\mathbf{y}_t$ is the unobserved latent process at time t. The parameters $\boldsymbol{\alpha}$, $\boldsymbol{\eta}$, $\mathbf{B}$, $\mathbf{Z}$, $\mathbf{T}$, and $\mathbf{P}$ are also sometimes allowed to depend on t.

Such a model is known as a multivariate autoregressive state-space (MARSS) model. The terms vector autoregressive state-space model or dynamic linear model are also used in various contexts.[12] MARSS models extend vector autoregression by allowing for measurement error. They also allow for the possibility of missing observations, as yt can be treated as

[11] More specifically, the observations are assumed to be drawn from independent multivariate normal distributions with expectation equal to a linear function of the patent process.

[12] Terminology varies depending on discipline and model use. Regardless of the name and interpretation of the parameters and latent process, the models are mathematically identical. The term dynamic linear model is commonly used when the latent process represents regression parameters, thus creating a model in which the regression parameters change with time. This distinction is essentially conceptual, with the state space usually capturing regression parameters and independent variables appearing in the mapping between the state space and the expected value of an observation rather than the other way around. The MONOCAR model introduced in this paper can also be used to capture dynamic regression parameters in a similar fashion to dynamic linear models.

unobserved at some points in time.[13]

MONOCAR, the model introduced in this paper, can be thought of as a type of state-space model but differs from multivariate state-space models and other existing state-space models. If a state-space model is defined as requiring an observation at time t to depend only on the state at time t, then MONOCAR cannot be treated as a state-space model in which the state at time t is the value of the latent process at time t because MONOCAR allows non-instantaneous observations that depend on all values of the latent process over a continuous interval of time.[14] However, a state-space representation of MONOCAR is possible using a representation in which the state-space differs from the space of the latent process.[15]

Most importantly, MONOCAR differs from MARSS models by treating the latent process as a continuous-time process and allowing an observation to be based on a sum or average over a continuous interval of time, including an interval that overlaps the time intervals of other observations.[16] This allows MONOCAR to handle temporally aggregated data even when such aggregation is inherent in the data, some or all of the time intervals are overlapping, and time is not naturally divided into discrete units. MONOCAR also addresses the critique of Pickup and Wlezien (2009) regarding the variance of observations from polling data (see section 3.4).

[13]While MARSS models are thus able to deal with irregularly spaced observations, they are not quite as flexible as continuous-time models in this respect.

[14]Despite the fact that the model treats non-instantaneous observations as depending on the value of the latent process at infinitely many point in time, estimation of the model does not require estimating the value of the latent process at all these points of time.

[15]This representation essentially forms the basis of the likelihood computation.

[16]Although MARSS models are discrete time, continuous-time state-space models have also been studied. MONOCAR differs primarily in allowing observations to be based on sums or averages over continuous intervals of time. This complicates estimation of either the latent process or the model parameters. In this case, one cannot simply apply a Kalman filter with the state space representing the latent process at a given point in time because a non-instantaneous observation depends on the values of the latent process at infinitely many points in time. Additionally, much of the work on continuous-time state-space models has not been concerned with inference about model parameters. Often, the goal is estimation of latent process under known model parameters, which is more relevant in engineering contexts, or producing point estimates of the model parameters.

G.2. *Continuous-time models*

A continuous-time process is a process that is indexed by continuous, rather than discrete, values of time. In contrast, a discrete-time process is indexed by discrete—usually integer—values of time. While a discrete process might exist at all integer values of time and, thus, be continuous in the sense of being an uninterrupted sequence, it is not continuous in the sense of a continuous-time process. Thus, a discrete process, $\mathbf{y}_t$, might be defined for all integers t but would not be defined at non-integer values, such as $\mathbf{y}_{\frac{1}{2}}$ or even $\mathbf{y}_{\sqrt{2}}$.

In many situations, we conceptualize time as continuous but use discrete models for simplicity. If, however, we are modeling a process that is truly continuous, this simplification comes at a cost of reduced accuracy. A continuous-time model allows us to avoid this simplification.

Within political science, the use of continuous time has been almost entirely limited to event history analysis, in which durations are often treated as a continuous quantity, and longitudinal network analysis (e.g., Snijders 2001; Berardo and Scholz 2010; Manger, Pickup, and Snijders 2012; Fischer et al. 2012), with a very small number of additional uses in formal models (e.g., Fearon 1994, 1998; Carpenter 2003).

Continuous-time processes and discrete-time processes are heavily related. Discrete-time processes can often be represented as continuous-time processes that are only observed at regularly spaced points in time.[17] In this sense, one can often think of many discrete-time models as special cases of continuous-time models. Indeed, some textbooks on time series have introduced time-series models in a general framework that included continuous-time models before focusing on discrete-time models (e.g., Fuller 1976).[18]

However, it is useful to draw a distinction between models of this type and models that

[17]For example, a first-order vector autoregressive process can be viewed as a first-order multivariate continuous-time autoregressive process observed at regularly spaced points in time so long as the eigenvalues of $\mathbf{B}$ are positive.

[18]This is especially true in earlier textbooks when the literatures on continuous-time models and discrete-time models had diverged less than today.

cannot be represented as a discrete-time model. The term continuous-time model is more usefully applied to this latter type of models because continuous-time representations are rarely used when a simple discrete-time representation exists and because doing so allows us to contrast discrete-time models with continuous-time models. The term will be used in that sense in this paper.

Continuous-time models are most commonly applied in precisely these circumstances where discrete-time representations are not possible. For example, as Jones (1985) writes, "[w]hen data are truly unequally spaced, not equally spaced with missing observations, continuous time models are necessary to represent the process." It is thus unsurprising that a common use for continuous-time models is modeling this sort of data.

Continuous-time autoregressive (CAR) processes have received almost no use in political science but, like state-space models, have received significant attention in other disciplines. They have played an important role in physics for over a century and have also seen use in chemistry, biology, and engineering (see, e.g., Hänggi and Marchesoni 2005). Over the last several decades, their use has become common in finance—as demonstrated by the ubiquitous Black-Scholes option-pricing model, and applications in econometrics have also grown (see, e.g., Brockwell 2001a).

Despite receiving a number of uses in econometrics (see, e.g., Bergstrom 1990), most applications of CAR processes and their extension, continuous-time autoregressive moving-average (CARMA) processes, have been for quite different reasons than the difficulties motivating this paper. Among other reasons, continuous-time models are often chosen because the data is very high frequency, because they provide a more accurate representation of many macroeconomic models, because the results do not depend on the unit of time chosen for observations, or because they can generate continuous-time forecasts (Bergstrom 1996).

Some econometric applications are relevant to some of the issues discussed in this paper. Several papers have applied CARMA models to irregularly spaced data (Jones 1981, 1985; Jones and Ackerson 1990, e.g.,), as is discussed in more detail in section G.3. Many other

applications do allow observations to be averages or sums over intervals of time, although generally assuming these observations to be non-overlapping and regularly spaced. Measurement error is also commonly allowed. Finally, recent work has focused on extensions to multivariate continuous-time autoregressive (MCARMA) processes (Marquardt and Stelzer 2007; Schlemm and Stelzer 2012).

G.3. *Approaches to irregularly spaced data*

A number of approaches to the analysis of mixed-frequency and irregularly spaced data have been developed in other disciplines, particularly economics and astronomy. Some, such as those dealing with spectral analysis or forecasting, are relevant for research questions that are uncommon in political science. Others—especially those developed in economics—may be quite applicable, although few have attracted use in political science to date.

One approach has been to extend methods for regularly spaced time series to allow for missing observations. This work includes tests for the presence of serial autocorrelation (Savin and White 1978; Dufour and Dagenais 1985; Robinson 1985; Shively 1993) as well as tests for unit roots (Ryan and Giles 1998), and methods for estimation (Wansbeek and Kapteyn 1985) in its presence. Related work has addressed missing values in panel data, including estimation when errors are not autocorrelated (Wansbeek and Kapteyn 1989; Baltagi and Chang 1994) and testing for autocorrelation and estimation in its presence (Baltagi and Wu 1999). A Bayesian extension of the VAR model has also been proposed by Chiu et al. (2011).

Continuous-time models are another approach to analyzing irregularly spaced data. These models are particularly advantageous when the irregular spacing arises because the data lack an underlying sampling interval rather than arising due to missing observations in otherwise regularly spaced data, as most other methodologies deal only with the latter case.

Although much of this work has been focused on the univariate case (Jones 1981, 1985; Jones and Tryon 1987; Jones and Ackerson 1990), some has also addressed the multivariate

case (Jones 1984; Jones and Boadi-Boateng 1991). Other work has extended these models to non-Gaussian error processes (Brockwell 2001b; Yang 2008) and threshold models (Brockwell and Hyndman 1992; Brockwell 1994). Outside of work with an explicitly continuous-time model, Erdogan et al. (2005) present an approach that is also applicable to univariate time series with irregularly spaced observations that lack an underlying sampling interval.

Other work has focused on mixed-frequency data.[19] Although mixed-frequency data pose a simpler problem than irregularly spaced data in that each series is still regularly spaced, they are also inherently multivariate. Methods for analyzing mixed-frequency data include the use of bridge equations to link series of different frequencies (Baffigi, Golinelli, and Parigi 2004; Diron 2008, e.g.,), interpolation methods (Lanning 1986; Angelini, Henry, and Marcellino 2006, e.g.,), a state-space approach known as mixed-frequency vector autoregression (MF-VAR; Zadrozny 1988; Mittnik and Zadrozny 2005), and the Mixed-Data Sampling (MIDAS) approach (Ghysels, Santa-Clara, and Valkanov 2004) and its extensions, including an autoregressive extension (AR-MIDAS; Clements and Galvão 2008) and a more general one (unrestricted MIDAS; Foroni, Marcellino, and Schumacher 2012). Extensions to mixed-frequency data in which some series end before other series, known as ragged-edge data, have also been discussed (e.g., Mariano and Murasawa 2010).

Finally, another literature, primarily in astronomy, deals with analyzing the frequency spectrum of irregularly spaced data (Barning 1963; Lomb 1976; Scargle 1982; Vityazev 1996; Broersen 2008; Lévy-Leduc, Moulines, and Roueff 2008, e.g.,). While this work has found some applications in fields such as biology (Ruf 2010; Van Dongen et al. 2010) and climatology (Schulz and Stattegger 1997; Matyasovszky 2013), frequency-domain models like spectral analysis have attracted little use in political science.[20] More importantly, frequency-domain models are generally relevant to different research questions than are time-domain models such as the methodology discussed in this paper (Van det Eijk and Weber 1987).

[19] For a detailed overview of this literature, see Foroni and Marcellino (2013).

[20] But see Aguiar-Conraria, Magalhães, and Soares (2012); Im, Cauley, and Sandler (1987); Gregory and Gallagher (2002); Sandler and Enders (2004) for examples of their use in political science and related fields and Van det Eijk and Weber (1987) for a discussion.

H. COMPLETE MONTE CARLO STUDIES

H.1. *Temporal aggregation*

Non-overlapping intervals To illustrate some of the consequences of temporal aggregation for VAR and MONOCAR models, consider the following Monte Carlo study. In each simulation, the data are generated by aggregating two continuous-time latent processes, y_1 and y_2, that are constructed so that y_2 affects y_1 but not the reverse.[21] These latent processes are then aggregated into averages over 1,000 successive time intervals of unit length.[22] I will focus on the coefficient on lagged values of the first series in the equation predicting the second series, denoted $B_{2,1}$, which might be used in testing the effect of the first series on the second series.

I vary the mean reversion rate of the two processes, described by $\Theta_{1,1}$ and $\Theta_{2,2}$, and the size of the effect of the second series on the first, described by $\Theta_{2,1}$.[23] Although $\Theta_{2,1} \leq 0$ in all simulations, indicating that increases in the second series tend to increase the first, similar results with bias in the opposite direction can be expected $\Theta_{2,1} > 0$.[24] In all simulations, however, $\Theta_{1,1} = 0$.

The results of the simulations show how temporal aggregation can lead us to misunderstand the causal relationship between series. Although, by assumption, the first series has no effect on the second series, the VAR estimates seems to suggest otherwise.

Figure A2 shows the distribution of $\hat{B}_{2,1}$ from VAR models.[25] In the first row of figure A2, neither series affects the other ($B_{1,2} = 0$). With the two processes entirely independent, VAR

[21] Specifically, underlying process is the multivariate Ornstein-Uhlenbeck process given by

$$d\mathbf{y}_t = \boldsymbol{\Theta}\left(\boldsymbol{\mu} - \mathbf{y}_t\right) dt + \boldsymbol{\Sigma}^{\frac{1}{2}} d\mathbf{W}_t,$$

with $\boldsymbol{\mu} = 0$, $\boldsymbol{\Sigma} = \mathbf{I}_2$, and $\boldsymbol{\Theta} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}$.

[22] That is, the aggregated observation i at time t, $\tilde{y}_{it}$, is defined by $\tilde{y}_{it} = \int_{t-1}^{t} y_i \, dt$ for all $t \in \{1, \ldots, 1000\}$.

[23] I vary $\Theta_{2,1}$ from 0 to 1 in increments of 0.25. I vary $\Theta_{1,1}$ and $\Theta_{2,2}$ from 0.25 to 1. I do not consider the case in which $\Theta_{1,1} = \Theta_{2,2} = 0$, which would create a non-stationary process subject to the unit root problem discussed in section 3.7.

[24] Note that changing the sign on $\Theta_{2,1}$ is equivalent to replacing y_2 by $\mu_2 - y_2$.

[25] The results are based on a 1,000 simulations per study.

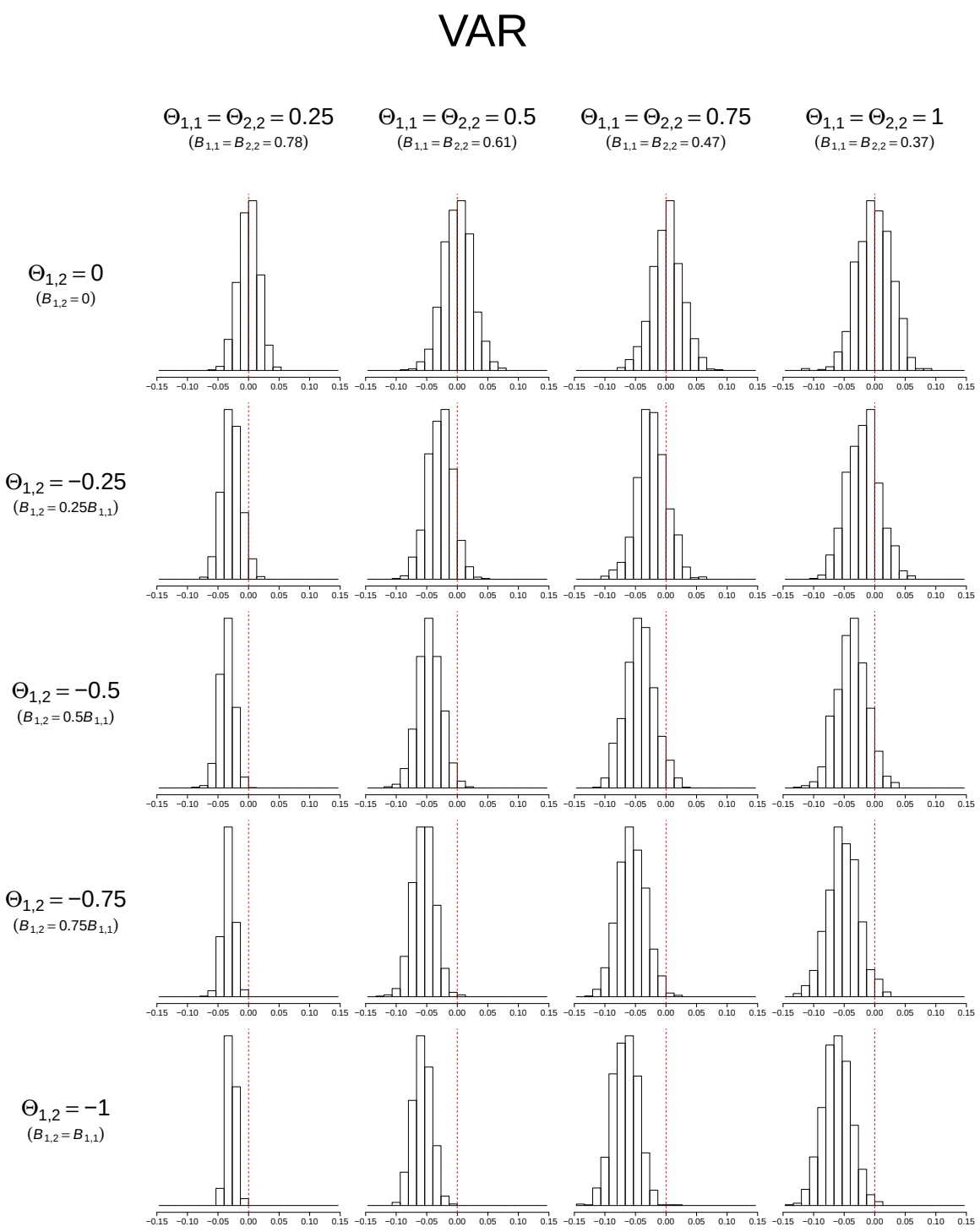
models show no bias in their estimates of $B_{2,1}$ regardless of the mean reversion rate of each series, which varies by column. However, as we move down the rows of the figure, the effect of the second series on the first is increased. Although the true value of $B_{2,1}$ remains zero, VAR estimates show increasing bias as we increase this effect. Moving across the columns of this figure, the bias appears to stay relatively constant as the mean revision rate of the two series changes.

We see similar patterns in the p-values generated by a VAR model. Figure A3 shows quantile-quantile plots comparing the empirical distribution of p-values to a uniform distribution. Because the null hypothesis is true in all cases, the p-values should follow a uniform distribution. If the p-values were uniform, the points in the quantile-quantile plots should lie very near the 45-degree line shown in the figure. When these points lie below the 45-degree, the model is rejecting the null hypothesis at a higher rate than the nominal level of a test. In the first row of figure A3, with neither series affecting the other, the p-values follow a uniform distribution. However, moving down the graph, the p-values show increasing deviations from the desired uniform distribution as the effect of the second series on the first increases.

Unlike VAR, MONOCAR estimates and p-values show no signs of bias. In figure A4, we see that the distribution of the estimates is centered almost perfectly around the true value of zero in all cases. Moreover, the points in the quantile-quantile plots shown in figure A5 lie very near the 45-degree line in all cases, indicating that the MONOCAR p-values follow a uniform distribution. This is exactly what we should see from an unbiased test of a true null hypothesis, as is the case here.

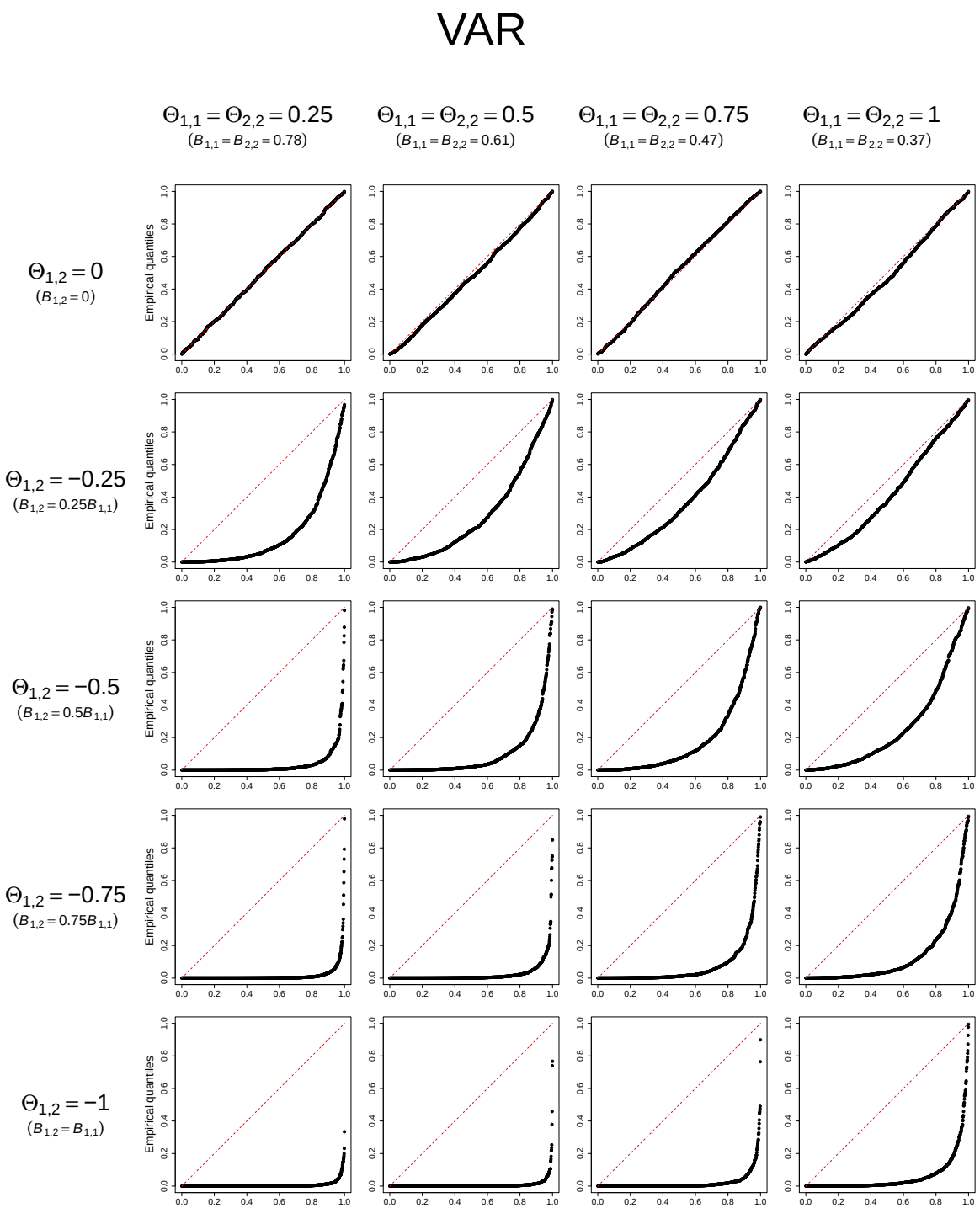
This study allows us to isolate the biases from temporal aggregation alone. Because the observations are regularly spaced, there is no need to drop or impute observations to use a VAR model. Although observations are regularly spaced, they represent averages over successive intervals of time (that is, temporal aggregates) rather than instantaneous measurements, as a VAR model implicitly assumes. The assumptions of a VAR model are not otherwise violated.

Figure A2: Histogram of distribution of estimates from VAR model for the temporal aggregation Monte Carlo study



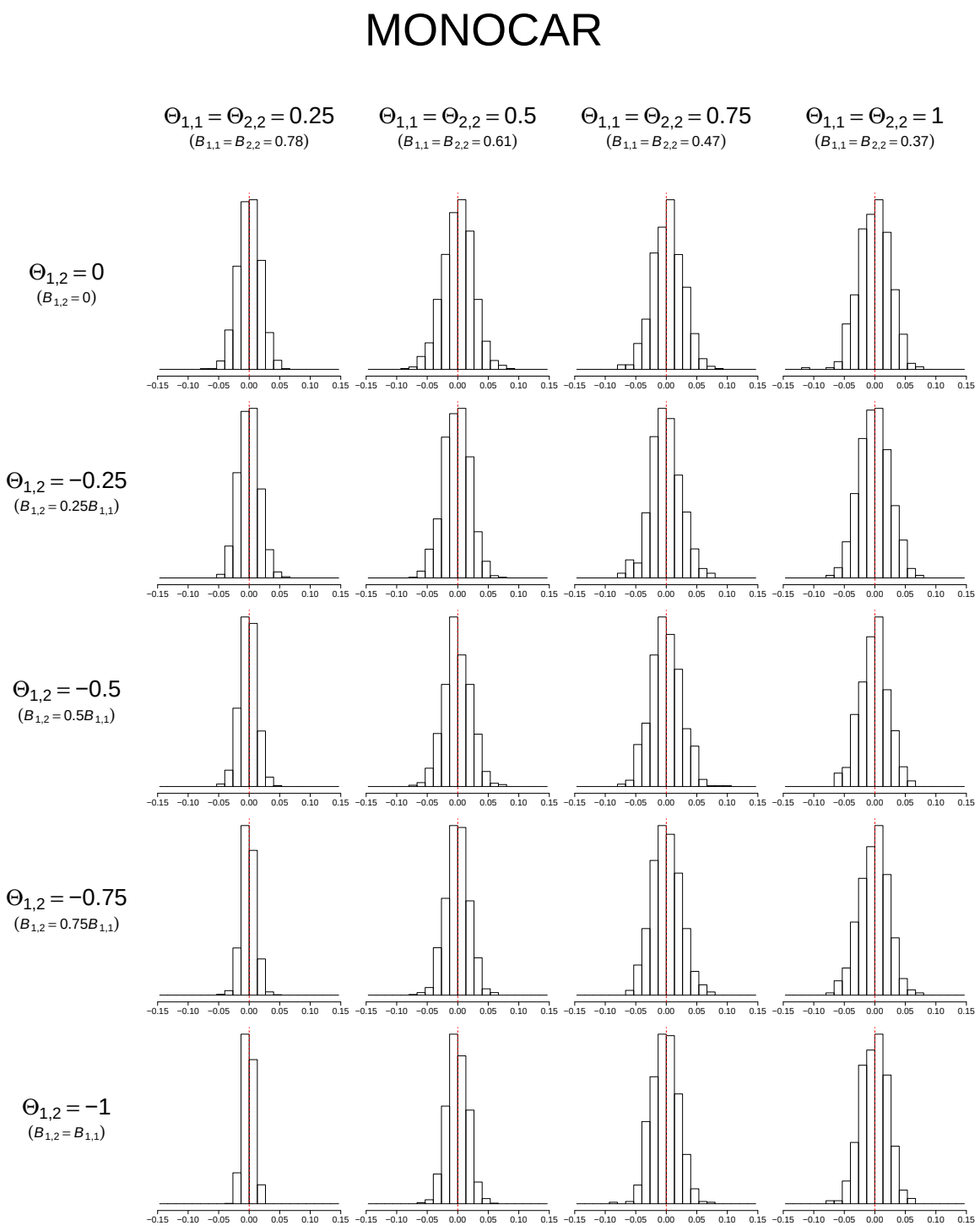
Note: 1,000 simulations per study. In each simulation, the data consist of 1,000 observations per series. The estimated parameter shown is $\mathbf{B}_{1,2}$, which equals zero under the data generating process. A dashed red line shows this true value of $\mathbf{B}_{1,2} = 0$ in each plot.

Figure A3: Quantile-quantile plot of p-values from VAR model for the temporal aggregation Monte Carlo study



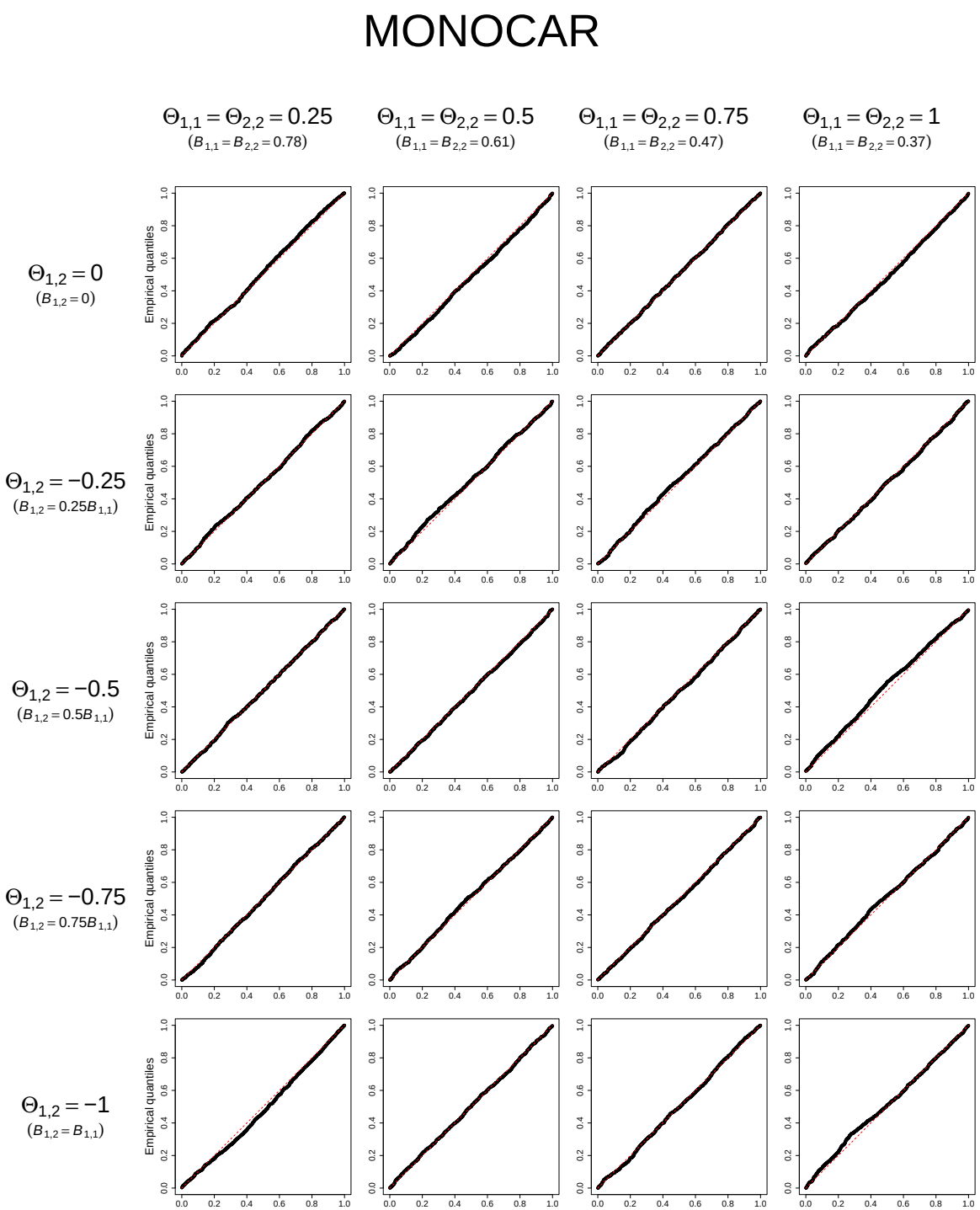
Note: 1,000 simulations per study. In each simulation, the data consist of 1,000 observations per series. The estimated parameter shown is $\mathbf{B}_{1,2}$, which equals zero under the data generating process. The p-values correspond to a test of the null hypothesis that $\mathbf{B}_{1,2} = 0$. Since the null hypothesis is true under the data generating process, valid p-values should follow a uniform distribution.

Figure A4: Histogram of distribution of estimates from MONOCAR model for the temporal aggregation Monte Carlo study



Note: 1,000 simulations per study. In each simulation, the data consist of 1,000 observations per series. MONOCAR estimates are transformed to use the same parameterization as the VAR model. The estimated parameter shown is $\mathbf{B}_{1,2}$, which equals zero under the data generating process. A dashed red line shows this true value of $\mathbf{B}_{1,2} = 0$ in each plot.

Figure A5: Quantile-quantile plot of p-values from MONOCAR model for the temporal aggregation Monte Carlo study



Note: 1,000 simulations per study. In each simulation, the data consist of 1,000 observations per series. MONOCAR estimates are transformed to use the same parameterization as the VAR model. The estimated parameter shown is $\mathbf{B}_{1,2}$, which equals zero under the data generating process. The p-values correspond to a test of the null hypothesis that $\mathbf{B}_{1,2} = 0$. Since the null hypothesis is true under the data generating process, valid p-values should follow a uniform distribution.

The bias in VAR estimates also changes little with the sample size. Graphs of Monte Carlo simulations using sample sizes of 250, 500, and 1,000 appear at the end of this section. While larger sample sizes unsurprisingly lead to lower variance in the estimates, the bias in each case is essentially unchanged as the sample size increases from 250 and 1,000. MONOCAR estimates show no apparent bias and a uniform distribution of p-values in all cases.

Since no measurement error or noise has been added to the observations, the failure of a VAR model to include this cannot explain the results. This also has implications for the multivariate autoregressive state-space model discussed in section 2.5. Given the lack of noise, a multivariate autoregressive state-space (MARSS) model becomes equivalent to this VAR model, as the observations must be equal to the values of the state space. Thus, a MARSS model can also exhibit temporal-aggregation bias, as in this case.

Thus, the use of a VAR model on temporally aggregated data risks biased estimates and unreliable inferences even if the data contain no other violations of the VAR assumptions. In contrast, MONOCAR appears to provide an effective solution to the problem.

Overlapping observations Overlapping observation intervals have the potential to increase the bias in a VAR model. So far, this study has only considered the effects of temporal aggregation for time intervals for each observation did not overlap. I now extend this study to include overlapping observations.

Observations are again assumed to occur at integer values of time. However, rather than assume that each observation is measured over an interval of length 1, observation times are allowed to extend past the start time of the subsequent observation. The length of each observation interval is varied from 1, implying no overlap between observations, to 3, implying that two-thirds of the time interval for one observation overlaps with the time interval for the next.

In order to simplify the presentation, only a single parameterization of the latent process[26]

[26]Specifically, I assume that $\mathbf{\Theta} = \begin{pmatrix} 1 & -\frac{1}{2} \\ 0 & 1 \end{pmatrix}$ and $\mathbf{Sigma} = \mathbf{I}$.

will be considered, although similar patterns to those observed in the first study can be expected if these parameters are varied.

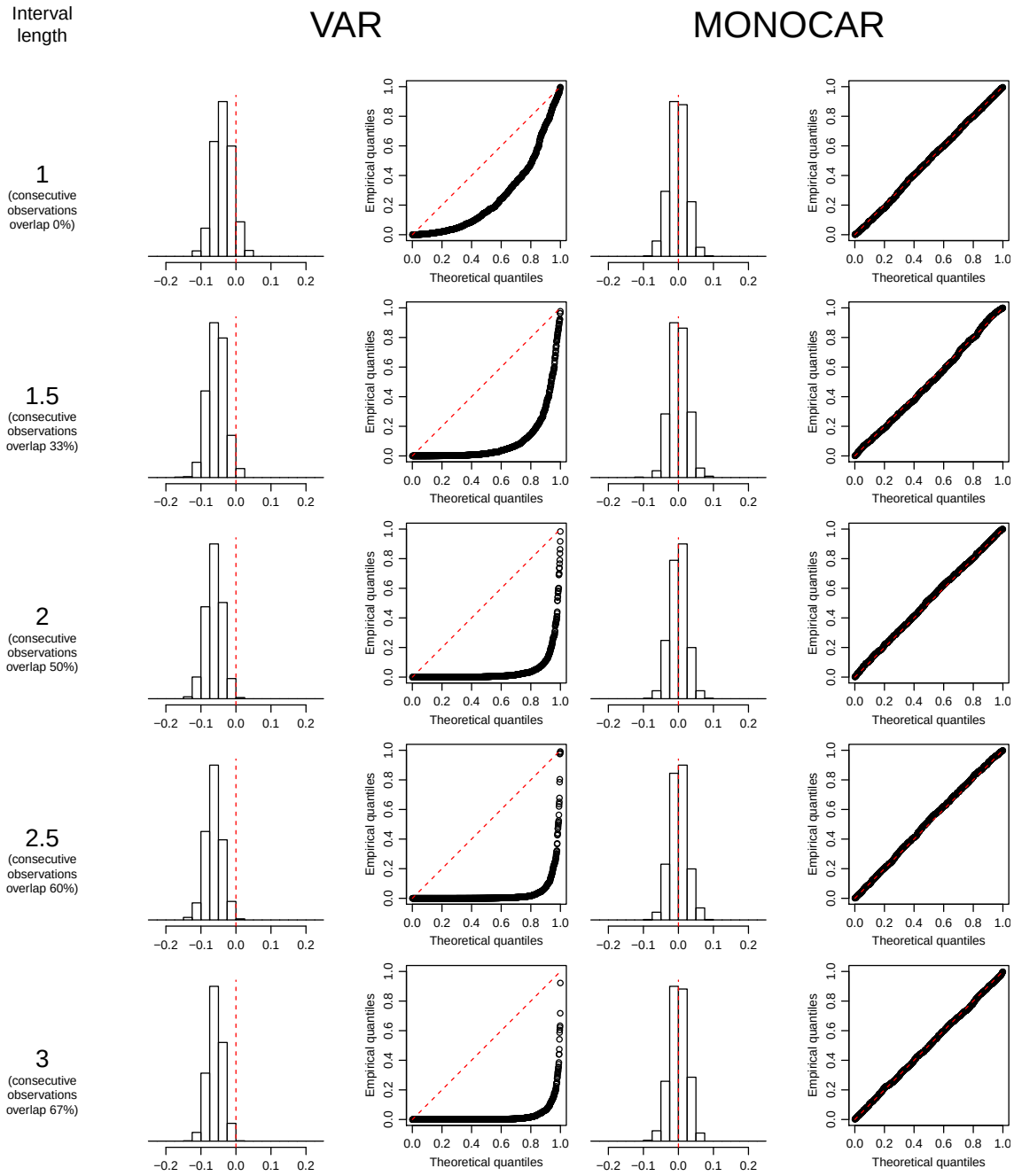
Figure A6 shows the distribution of estimates and quantile-quantile plot of p-values against a standard uniform. For the VAR models, we again see the negative bias in the estimates seen in the previous study. As the amount of overlap between observation intervals increases, the size of this bias increases. A similar pattern is observed in the p-values. Although the estimates and p-values are biased even when no overlap is present, a 50% overlap between observations creates a situation in which virtually all estimates are negative and the null hypothesis is erroneously rejected in most cases.

MONOCAR models, on the other hand, exhibit no bias regardless of the amount of overlap between observations. Similarly, the quantile-quantile plots show that the p-values from the MONOCAR models follow the desired uniform distribution that should be expected under the null hypothesis. Thus, while temporal aggregation poses an increasing problem for VAR models as the amount of overlap between time intervals increases, MONOCAR continues to produce reliable estimates and inference.

Overlapping observations with measurement error I next extent this study to include measurement error, creating a case in which MARSS models differ from VAR models. In most respects, the underlying data generating process is unchanged. The underlying latent processes considered and the length of the intervals are allowed to vary in the same fashion as in the previous simulations. However, rather than assume the observations occur without measurement error, the standard deviation of the measurement error, $\psi_{i,j}$, is assumed to take on one of four values—0.01, 0.05, 0.1, or 0.25—and is treated as known. With this change, a first-order MARSS model is no longer identical to a first-order VAR model.

Figures A7 and A9 show the distribution of estimates of these models using MARSS and MONOCAR, respectively. As in the previous study, MARSS models exhibit bias that increases as the amount of overlap between observations increases. Although the MARSS models take into account the measurement error, their performance is only marginally better

Figure A6: Histogram of distribution of estimates and quantile-quantile plot of p-values from VAR and MONOCAR models for the overlapping observation Monte Carlo study



Note: 1,000 simulations per study. In each simulation, the data consist of 1,000 observations per series. MONOCAR estimates are transformed to use the same parameterization as the VAR model. The estimated parameter shown is $\mathbf{B}_{1,2}$, which equals zero under the data generating process. The p-values correspond to a test of the null hypothesis that $\mathbf{B}_{1,2} = 0$. Since the null hypothesis is true under the data generating process, valid p-values should follow a uniform distribution.

than the VAR models. In contrast, the MONOCAR estimates exhibit no apparent bias.

The distribution of p-values—shown in figures A8 and A10—also demonstrate a similar pattern. MARSS models show bias in favor of rejecting the null hypothesis that increases with the amount of overlap between observations, although this bias decreases as the size of the measurement error increases. MONOCAR models again show a uniform distribution of p-values in all cases.

H.2. *Imputation*

While the previous Monte Carlo study dealt with bias from temporal aggregation, I now turn to a different source of bias—imputation. One common solution to irregularly spaced polls is to use WCALC or Samplemiser to create regularly spaced data, which are then used in VAR or other time-series models. This study will test this approach on simulated, irregularly spaced data and compare the resulting estimates and p-values with those from MONOCAR.

Imputation does not apply to all models that have been used to study irregularly spaced polls. Models such as MARSS models can deal effectively with irregularly spaced data occurring in discrete-time, although other problems such as temporal aggregation and overlapping observations seen in the previous studies can still cause issues. In this study, MARSS models can be expected to perform similarly to MONOCAR.[27] However, because the WCALC-imputation approach is often taken with polling data, it is worth analyzing the bias that results from this approach.

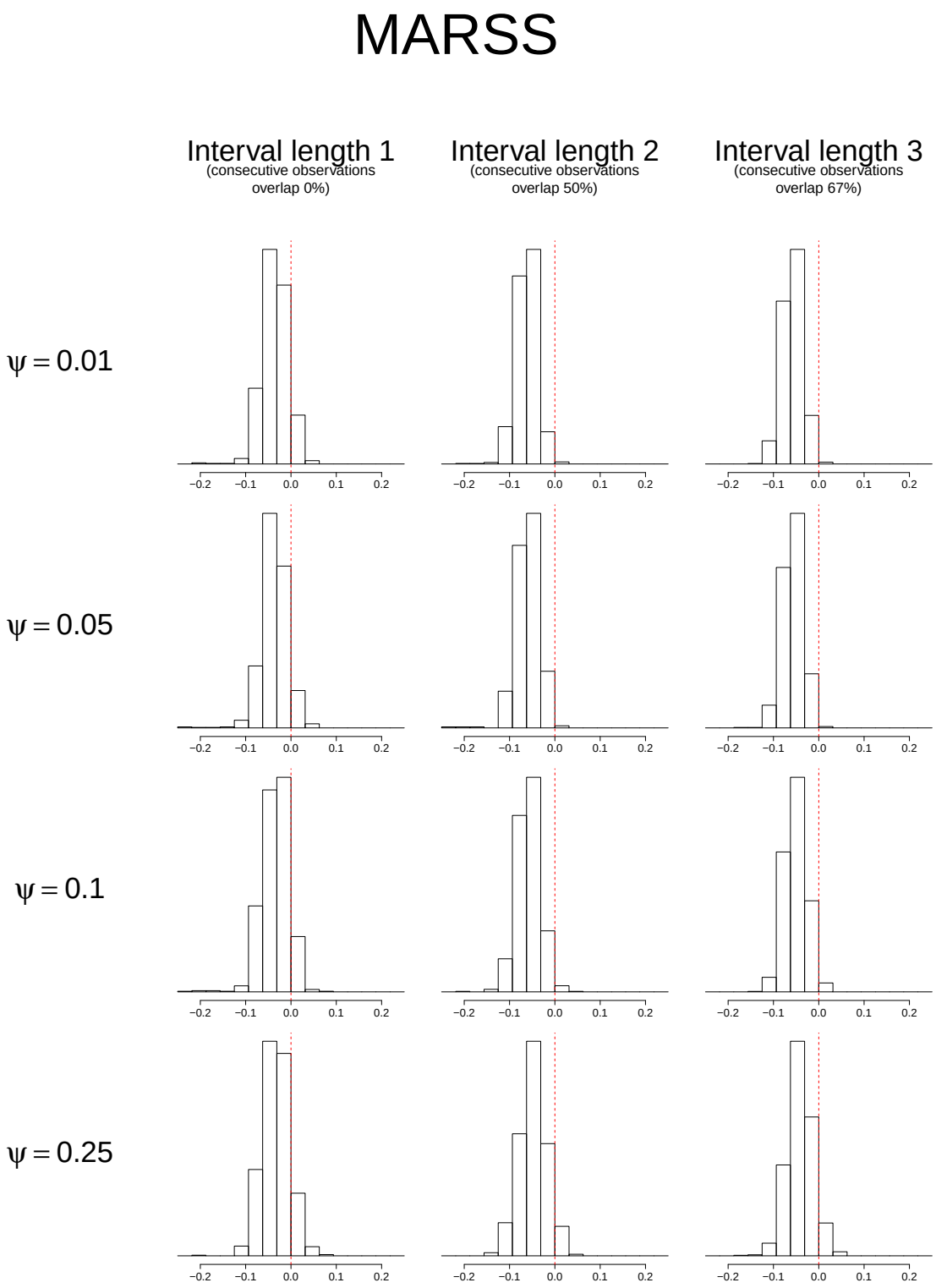
The simulated latent processes are identical to those used in the first Monte Carlo study.[28] This means that y_1 does not affect y_2 under the true data generating process. I will again focus on the estimated effects of y_1 on y_2 and p-values for the null hypothesis of no effect.

However, the observation process is different from the previous study. Observations are

[27] Because the data generating process in this study assumes instantaneous observations with spacing that are always multiples of a day (i.e., 24 hours), the MARSS and MONOCAR models are identical here except in their parameterizations.

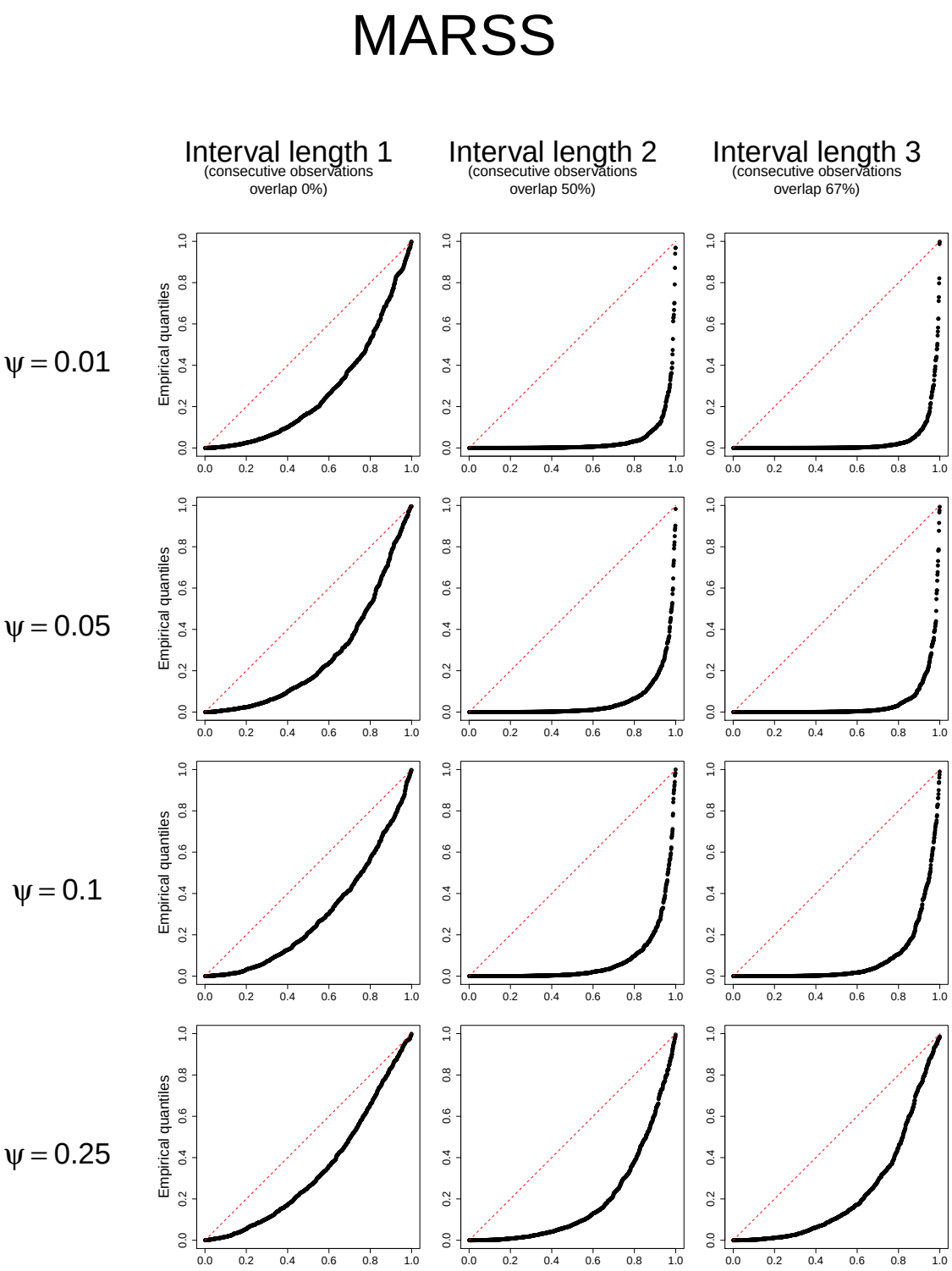
[28] See footnote 21 for details.

Figure A7: Histogram of distribution of estimates from MARSS model for the overlapping observation with measurement error Monte Carlo study



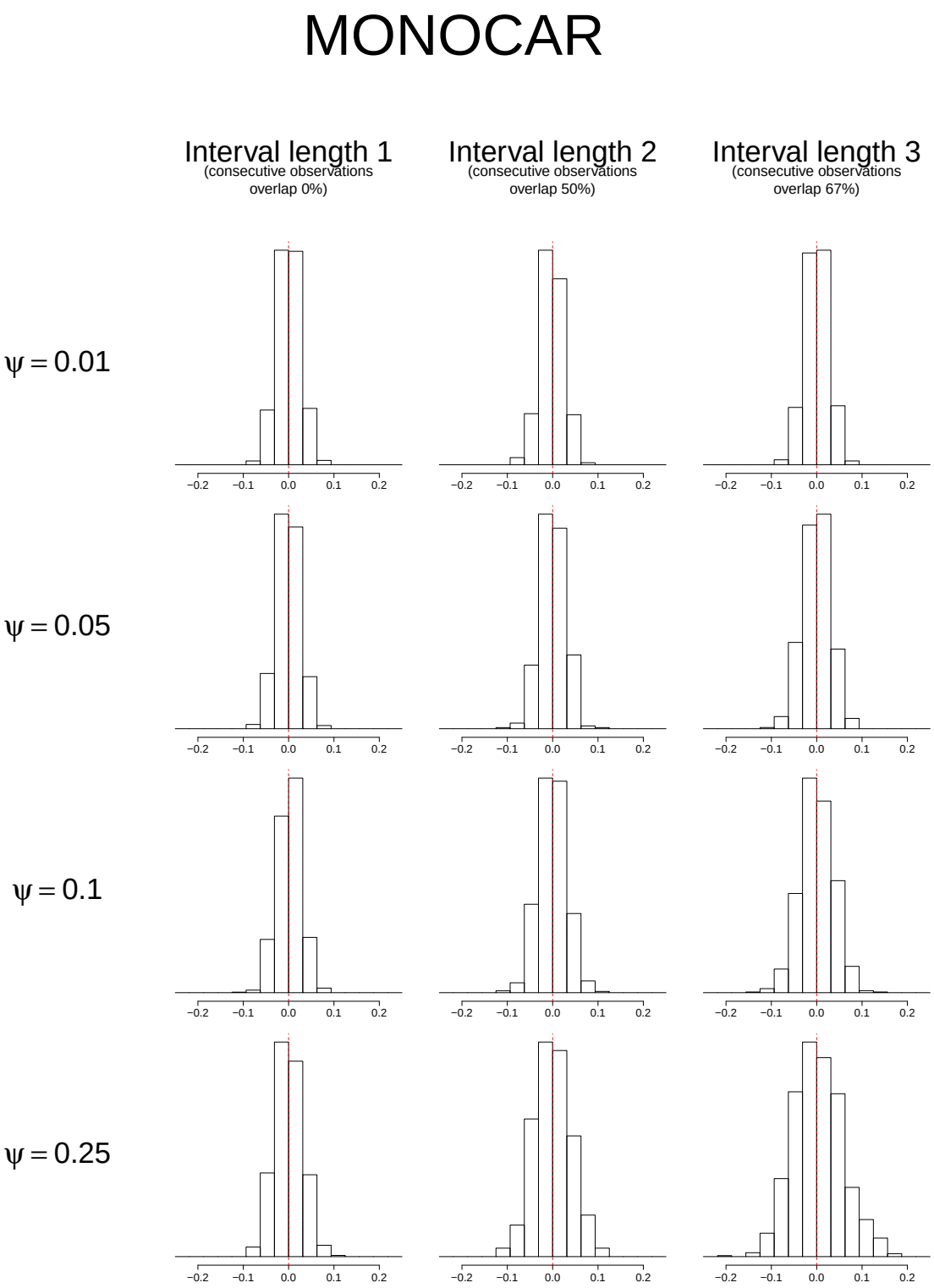
Note: 1,000 simulations per study. In each simulation, the data consist of 1,000 observations per series. ψ is the standard deviation of the measurement error for each observation. The estimated parameter shown is $\mathbf{B}_{1,2}$, which equals zero under the data generating process. A dashed red line shows this true value of $\mathbf{B}_{1,2} = 0$ in each plot.

Figure A8: Quantile-quantile plot of p-values from MARSS model for the overlapping observation with measurement error Monte Carlo study



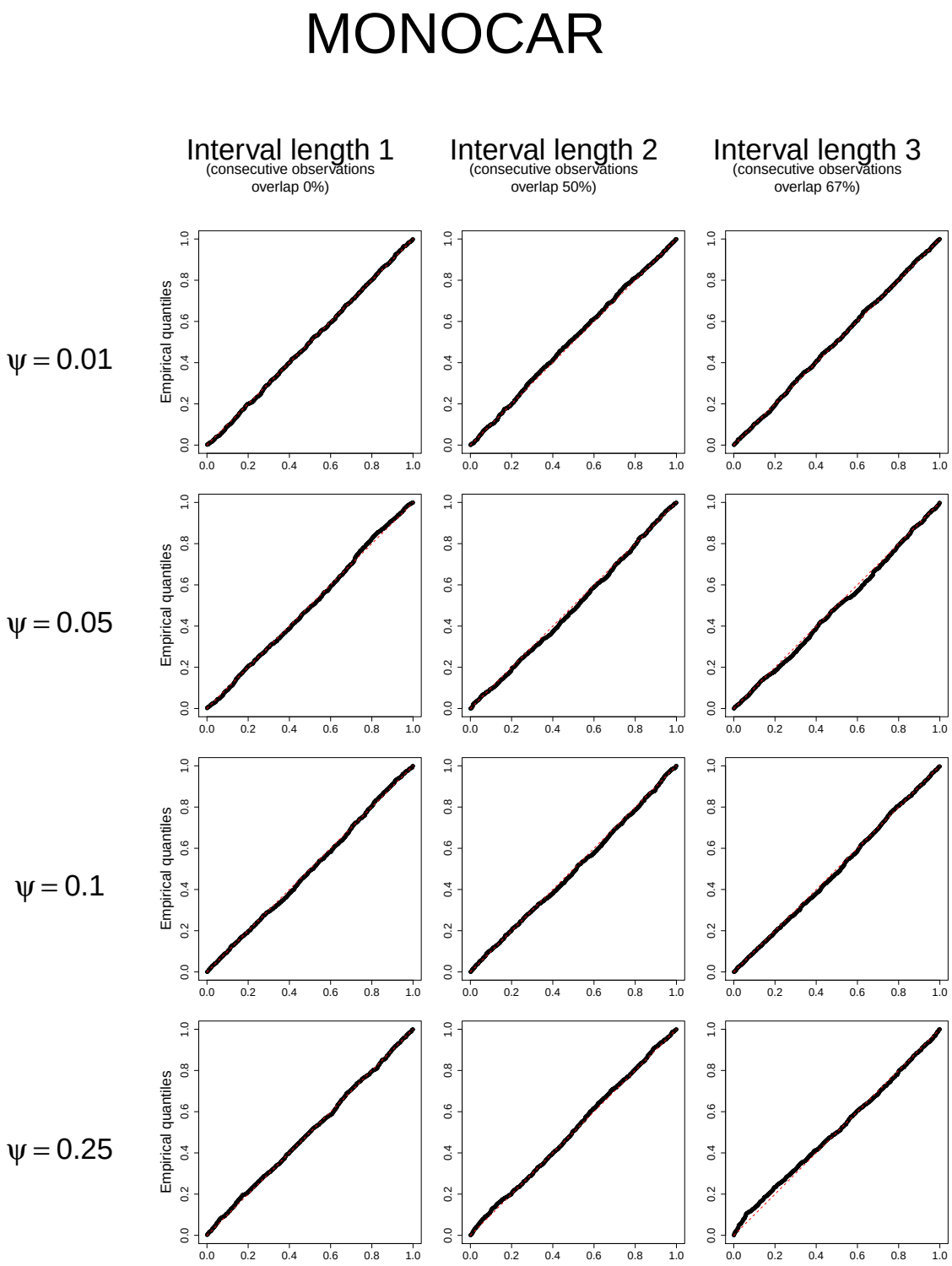
Note: 1,000 simulations per study. In each simulation, the data consist of 1,000 observations per series. ψ is the standard deviation of the measurement error for each observation. The estimated parameter shown is $\mathbf{B}_{1,2}$, which equals zero under the data generating process. The p-values correspond to a test of the null hypothesis that $\mathbf{B}_{1,2} = 0$. Since the null hypothesis is true under the data generating process, valid p-values should follow a uniform distribution.

Figure A9: Histogram of distribution of estimates from MONOCAR model for the overlapping observation with measurement error Monte Carlo study



Note: 1,000 simulations per study. In each simulation, the data consist of 1,000 observations per series. ψ is the standard deviation of the measurement error for each observation. MONOCAR estimates are transformed to use the same parameterization as the VAR model. The estimated parameter shown is $\mathbf{B}_{1,2}$, which equals zero under the data generating process. A dashed red line shows this true value of $\mathbf{B}_{1,2} = 0$ in each plot.

Figure A10: Quantile-quantile plot of p-values from MONOCAR model for the overlapping observation with measurement error Monte Carlo study



Note: 1,000 simulations per study. In each simulation, the data consist of 1,000 observations per series. ψ is the standard deviation of the measurement error for each observation. MONOCAR estimates are transformed to use the same parameterization as the VAR model. The estimated parameter shown is $\mathbf{B}_{1,2}$, which equals zero under the data generating process. The p-values correspond to a test of the null hypothesis that $\mathbf{B}_{1,2} = 0$. Since the null hypothesis is true under the data generating process, valid p-values should follow a uniform distribution.

instantaneous measures of the latent process in this study. However, the observations of the first series are not regularly spaced and both series include measurement error. Thus, temporal aggregation is not a problem with the simulated data, but irregular spacing and noise are.

The simulated data cover a 40-year period. Observations of the first series, which can be thought of as simulated polls, are collected at random points over this period.[29] Observations of the second series are regularly spaced and occur at the midpoint of each month. I then added measurement error to observations from both series.[30]

I analyze each set of simulated data using a typical process used with polling data. I first apply WCALC to the observations of the first series. To match the second series, I use monthly observation units. The result is imputed monthly observations for the first series which are paired with the corresponding monthly observations of the second series. I then estimate a VAR model on this data. I repeat this process using Samplemiser in place of WCALC.[31] Finally, I estimate a MONOCAR model on the original simulated data.[32]

The estimates and p-values from WCALC and VAR show clear evidence of bias. Figure A11 shows the distribution of estimates as we vary the parameters of the latent process. In the first row, the two latent processes are independent. As a result, the estimates are unsurprisingly unbiased. However, moving down the rows shows the bias as we vary the effect of the second series on the first. As this effect increases, we see increased bias in the estimated effect of the first series on the second. We can also see that the bias decreases as the mean reversion rate increases. That is, when the processes are more autocorrelated, we see greater bias in the estimates.

Similar issues arise with the p-values from WCALC and VAR. Quantile-quantile plots

[29]The observation times are independent and equally likely to fall on any data in the period. Unlike real polls, these observations are measured at a single moment in time.

[30]The measurement error consists of independent draws from a normal distribution with a mean of zero and standard deviation of 0.01 that are added to each observation.

[31]Samplemiser estimates are computed using a reimplementation of the approach described in Green, Gerber, and Boef (1999) with the MARSS library in R.

[32]That is, I use the raw data, not the imputed values from WCALC or Samplemiser, as MONOCAR does not require regularly spaced observations.

against a uniform distribution appear in figure A12. In the first column, in which the two latent processes are independent, p-values follow a uniform distribution. However, large deviations from a uniform distribution are visible when the second series affects the first. The p-values increasingly deviate from a uniform distribution as effect of the second series on the first increases and as the mean reversion rate decreases.

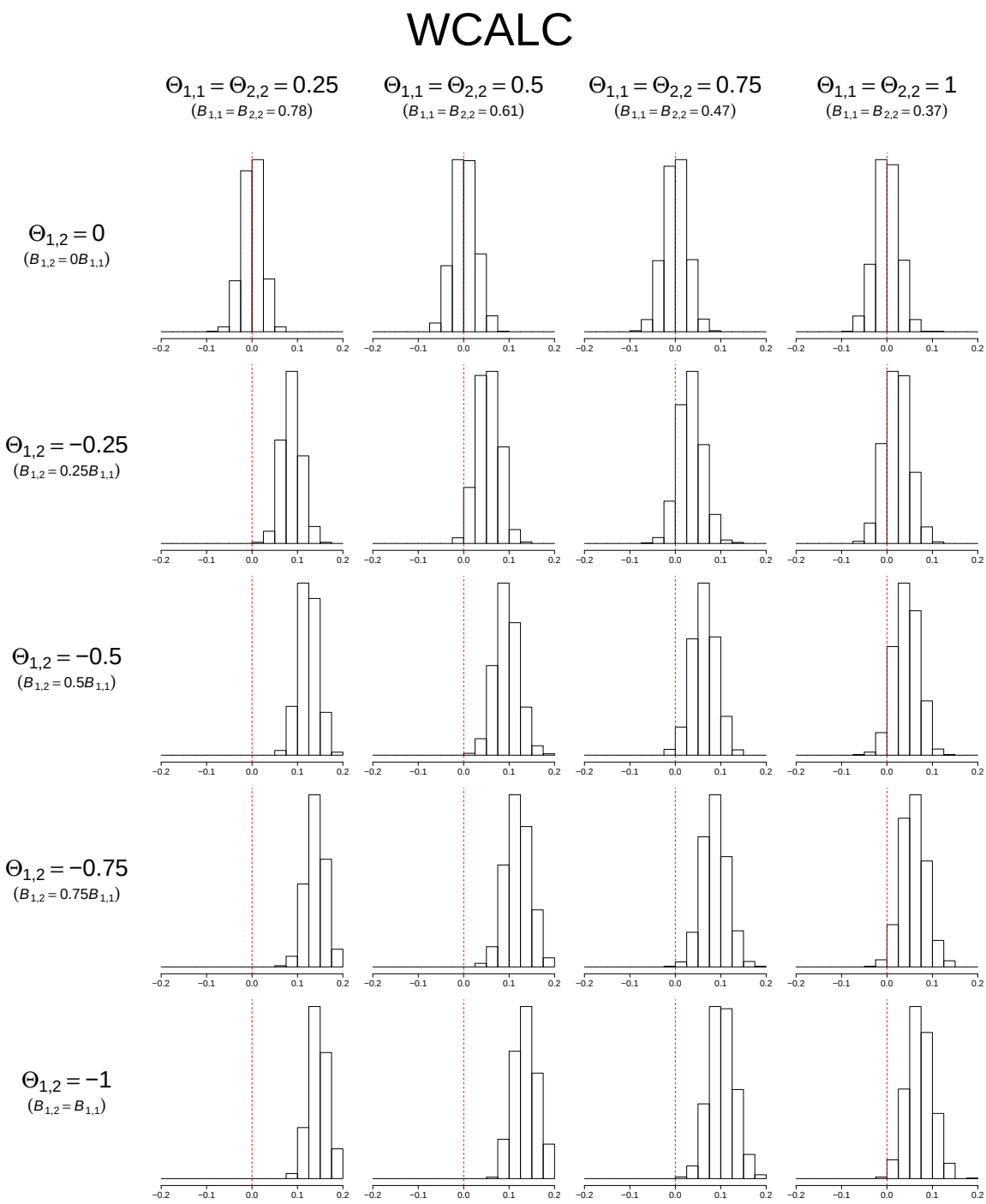
The results are very similar using Samplemiser in place of WCALC. The distribution of estimates using Samplemiser can be seen in figure A13 and quantile-quantile plots for the p-values can be seen in figure A14.

In contrast, MONOCAR again shows no signs of bias in either estimation or inference. Figure A15 shows the distribution of estimates, which appear centered around the true value of zero in all cases. Moreover, in the quantile-quantile plots that appear in figure A16, the p-values follow a uniform distribution, as the points in the graph lie on a 45-degree line, in all cases. Since the null hypothesis is true by assumption, this is precisely what we should observe from an unbiased test.

Thus, the common practice of using estimates of public opinion from WCALC or Samplemiser as data in a time-series analysis can be unreliable. Even if this process avoids temporal-aggregation bias, it does so at the cost of creating a new source of bias. However, by avoiding the need to aggregate or impute data, MONOCAR can produce unbiased estimates and inference.

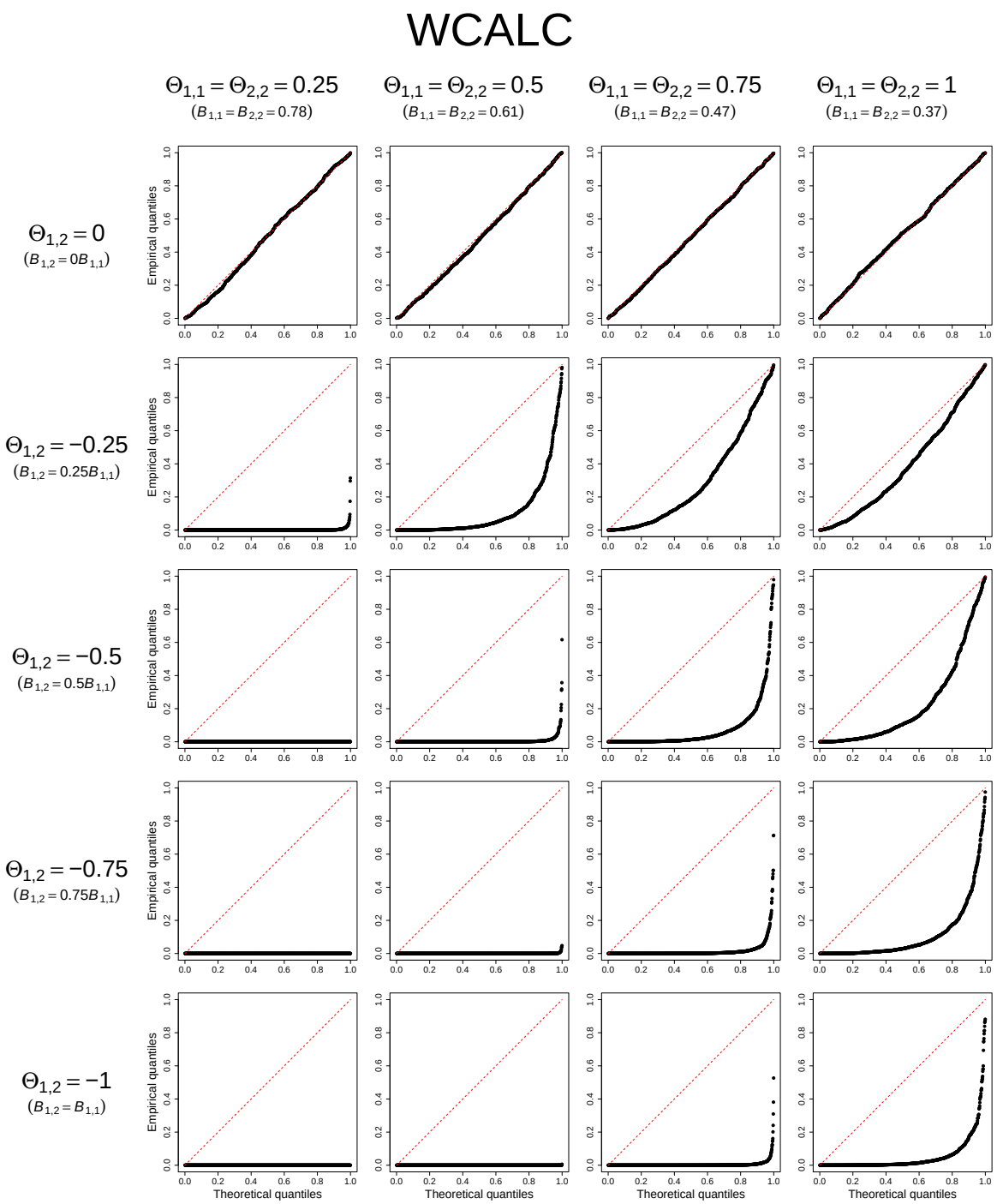
The bias illustrated in this study does not mean that WCALC, Samplemiser, or other such techniques are unreliable for other purposes. As estimators of the latent series themselves, they often perform quite well—including on the simulated data in this study. Rather, it is the use of estimated latent series as data that leads to the biases seen here.

Figure A11: Histogram of distribution of estimates from WCALC and VAR models in imputation Monte Carlo study



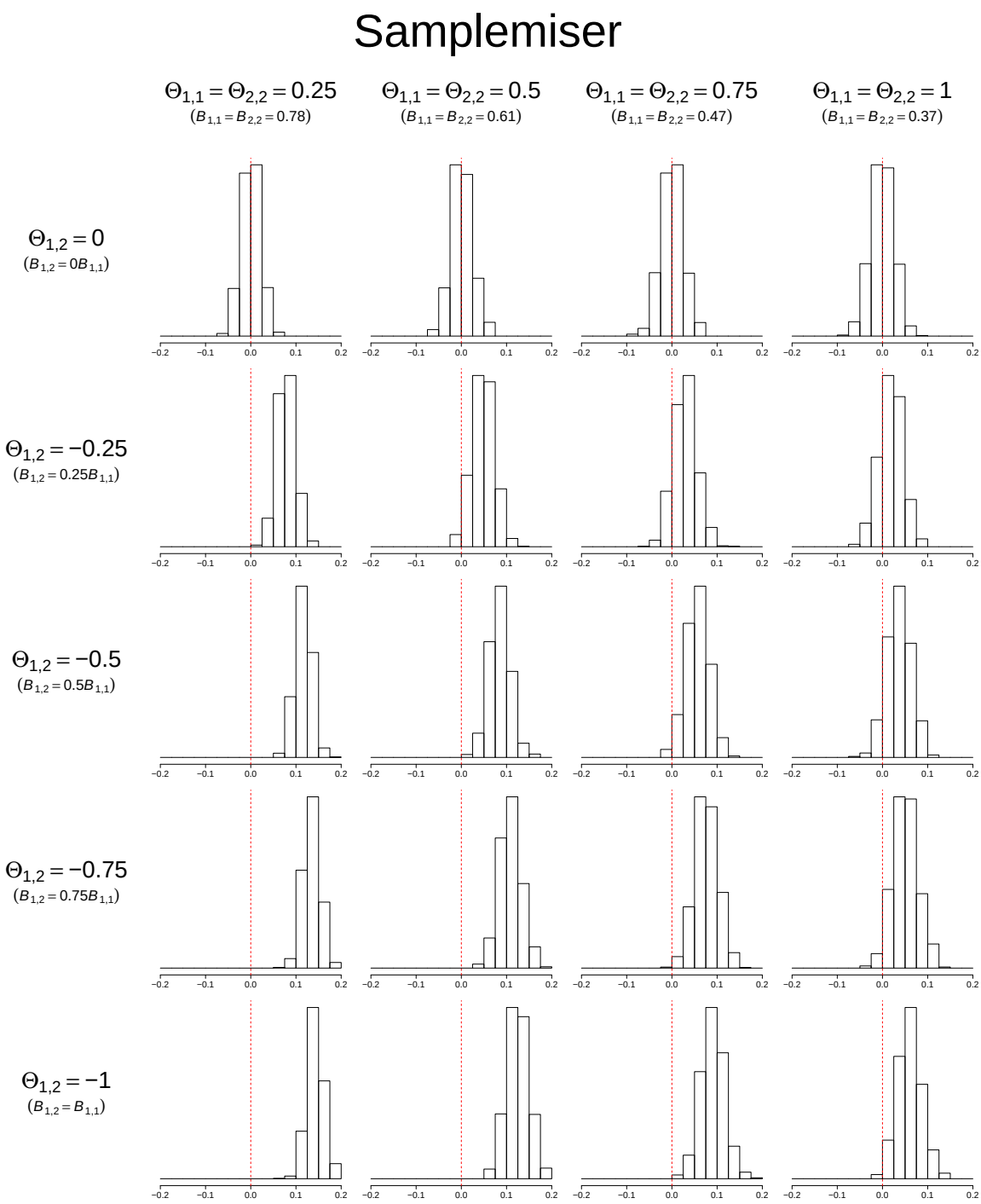
Note: 1,000 simulations per study. In each simulation, the first series consists of 160 randomly spaced instantaneous polls (averaging four per year) and the second series consists of monthly instantaneous observations. The estimated parameter shown is $\mathbf{B}_{1,2}$, which equals zero under the data generating process. A dashed red line shows this true value of $\mathbf{B}_{1,2} = 0$ in each plot.

Figure A12: Quantile-quantile plot of p-values from WCALC and VAR models in imputation Monte Carlo study



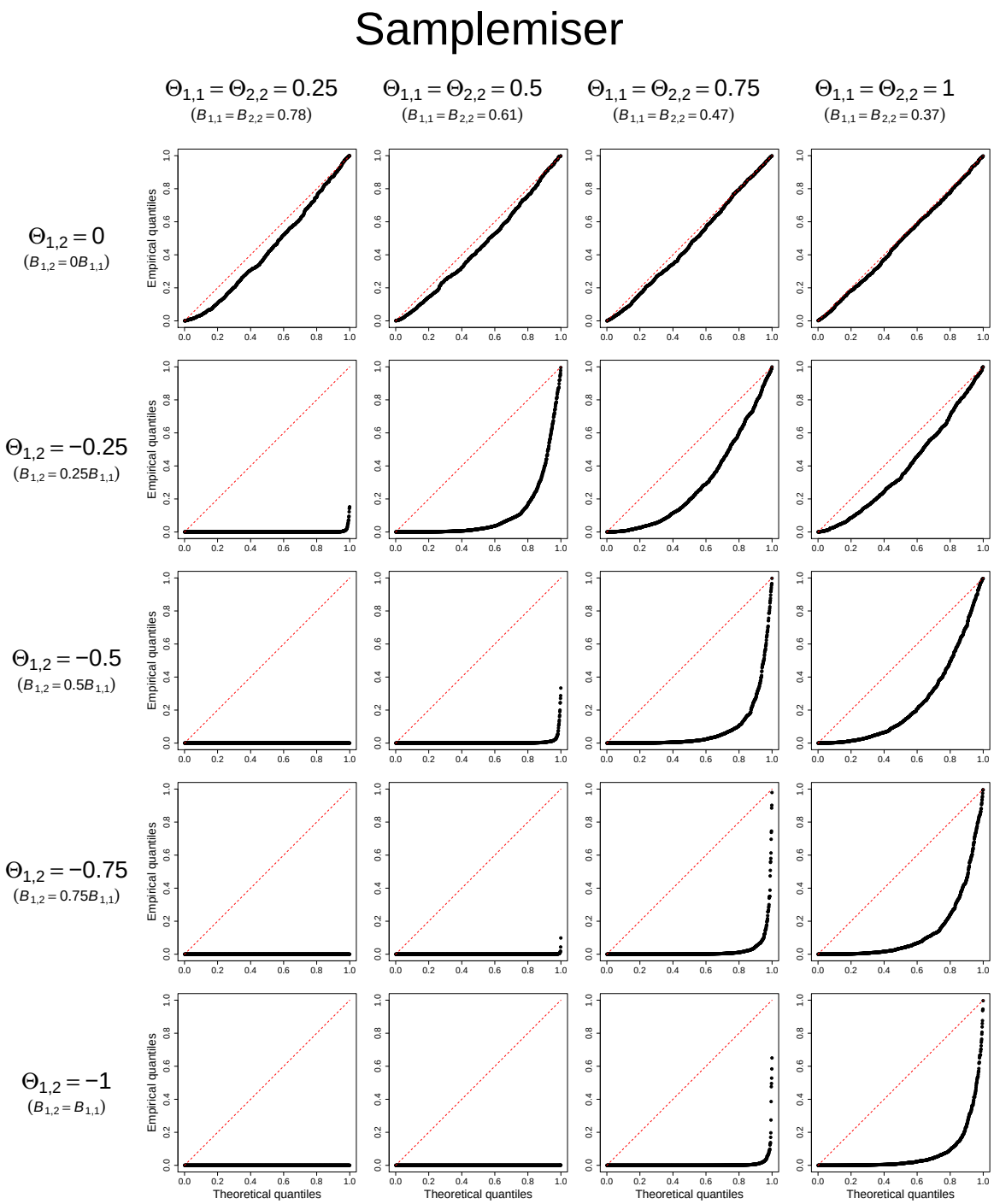
Note: 1,000 simulations per study. In each simulation, the first series consists of 160 randomly spaced instantaneous polls (averaging four per year) and the second series consists of monthly instantaneous observations. The estimated parameter shown is $\mathbf{B}_{1,2}$, which equals zero under the data generating process. The p-values correspond to a test of the null hypothesis that $\mathbf{B}_{1,2} = 0$. Since the null hypothesis is true under the data generating process, valid p-values should follow a uniform distribution.

Figure A13: Histogram of distribution of estimates from Samplemiser and VAR models in imputation Monte Carlo study



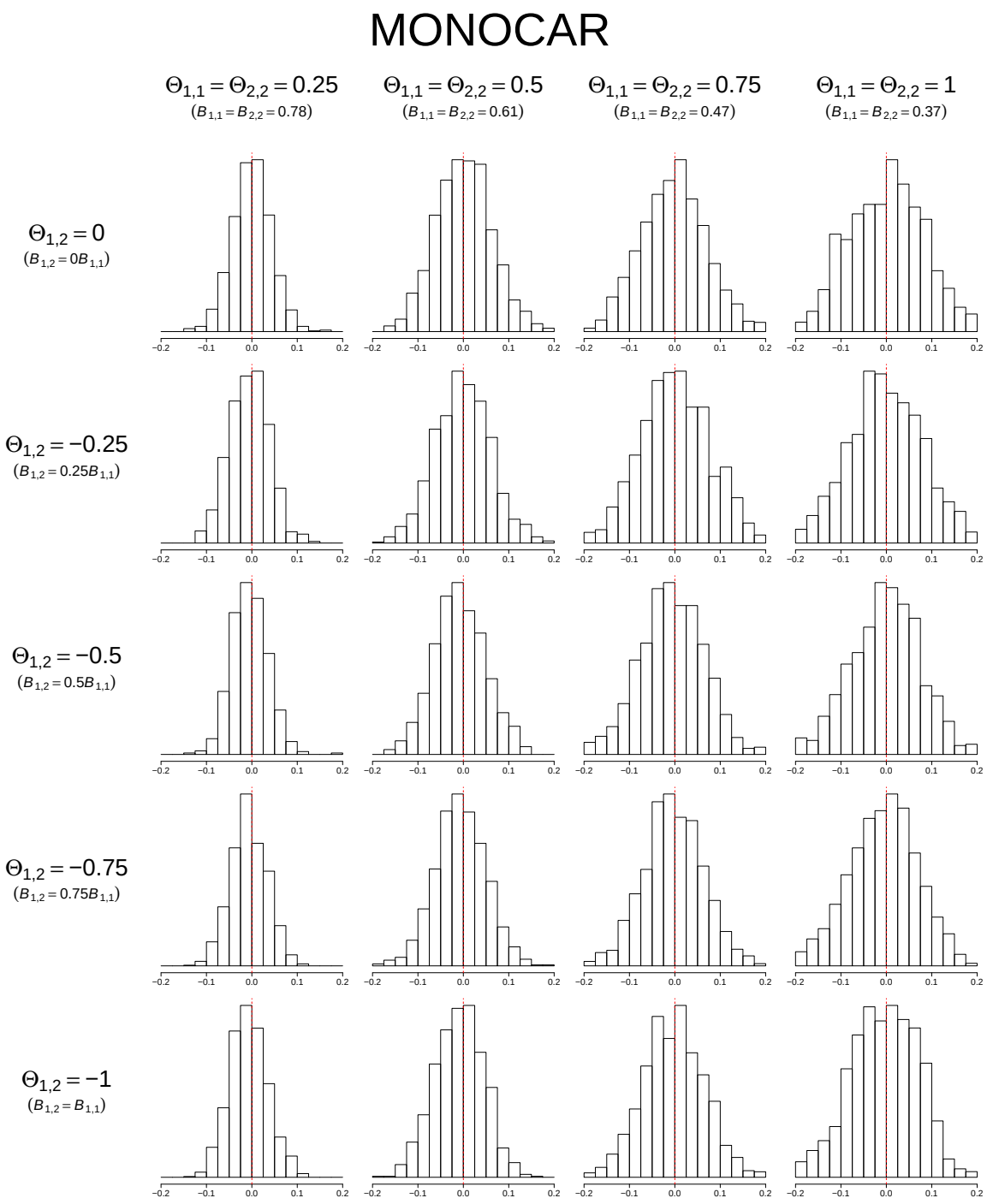
Note: 1,000 simulations per study. In each simulation, the first series consists of 160 randomly spaced instantaneous polls (averaging four per year) and the second series consists of monthly instantaneous observations. The estimated parameter shown is $\mathbf{B}_{1,2}$, which equals zero under the data generating process. A dashed red line shows this true value of $\mathbf{B}_{1,2} = 0$ in each plot.

Figure A14: Quantile-quantile plot of p-values from Samplemiser and VAR models in imputation Monte Carlo study



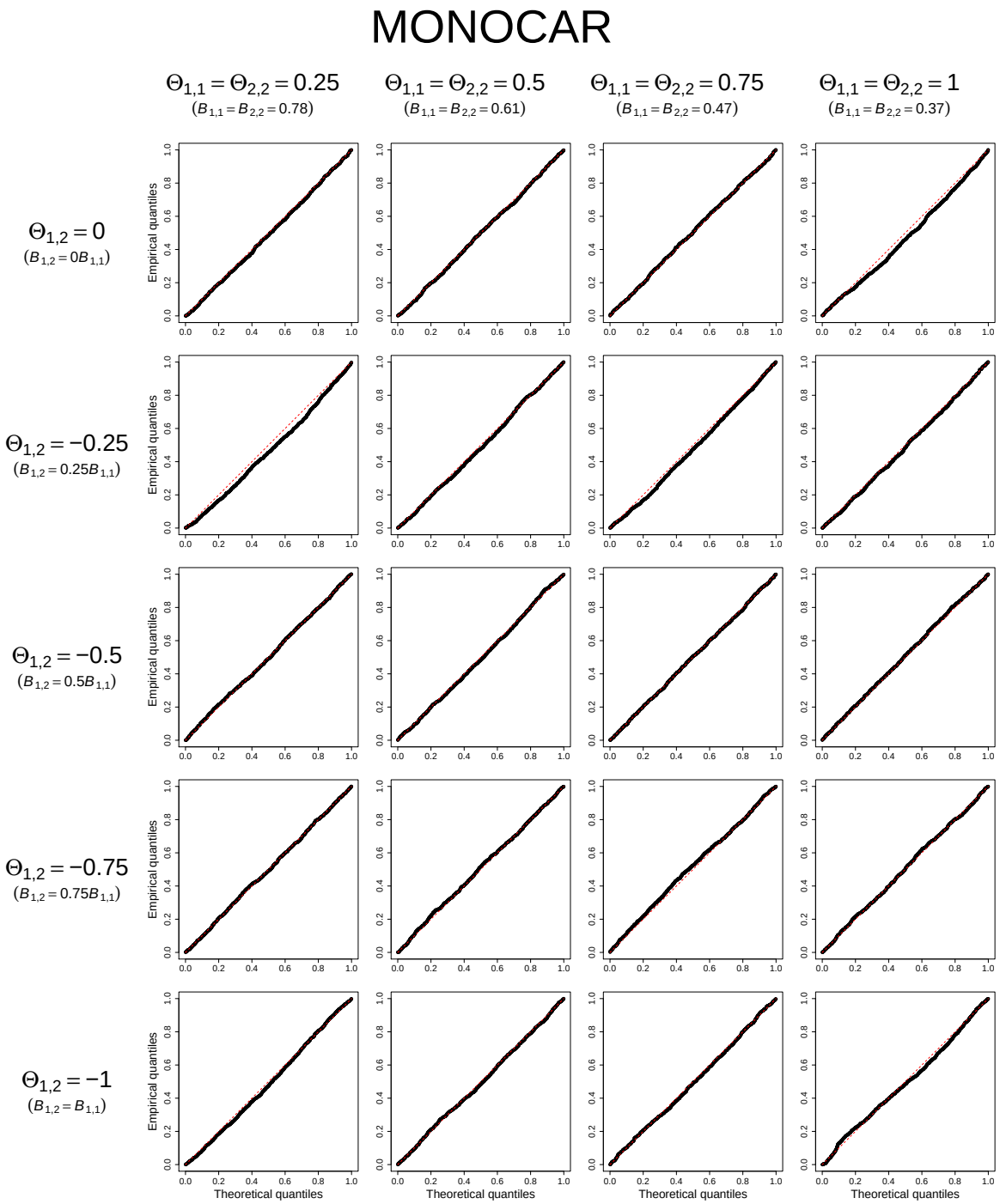
Note: 1,000 simulations per study. In each simulation, the first series consists of 160 randomly spaced instantaneous polls (averaging four per year) and the second series consists of monthly instantaneous observations. The estimated parameter shown is $\mathbf{B}_{1,2}$, which equals zero under the data generating process. The p-values correspond to a test of the null hypothesis that $\mathbf{B}_{1,2} = 0$. Since the null hypothesis is true under the data generating process, valid p-values should follow a uniform distribution.

Figure A15: Histogram of distribution of estimates from MONOCAR models in imputation Monte Carlo study



Note: 1,000 simulations per study. In each simulation, the first series consists of 160 randomly spaced instantaneous polls (averaging four per year) and the second series consists of monthly instantaneous observations. MONOCAR estimates are transformed to use the same parameterization as the VAR model. The estimated parameter shown is $\mathbf{B}_{1,2}$, which equals zero under the data generating process. A dashed red line shows this true value of $\mathbf{B}_{1,2} = 0$ in each plot.

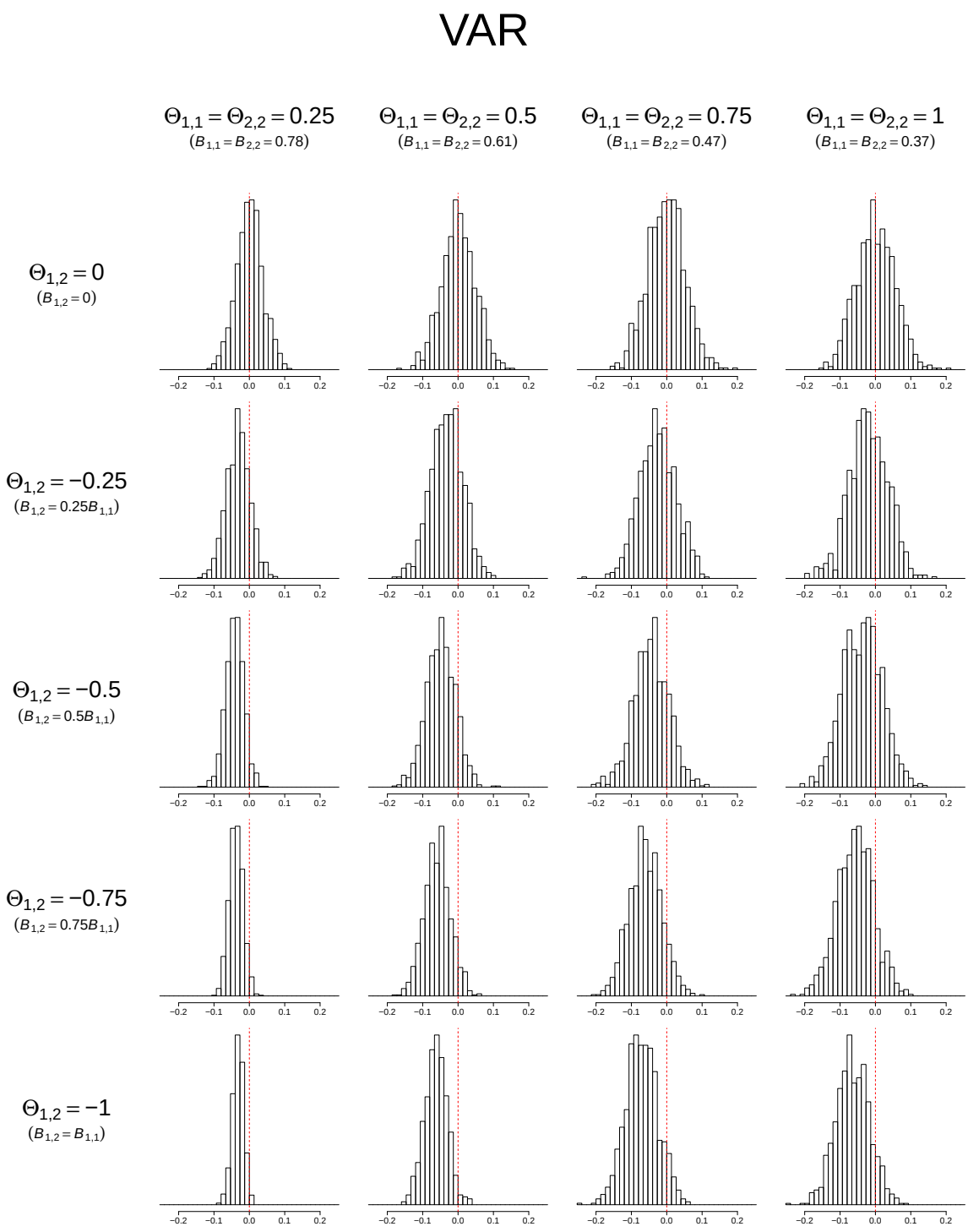
Figure A16: Quantile-quantile plot of p-values from MONOCAR models in imputation Monte Carlo study



Note: 1,000 simulations per study. In each simulation, the first series consists of 160 randomly spaced instantaneous polls (averaging four per year) and the second series consists of monthly instantaneous observations. MONOCAR estimates are transformed to use the same parameterization as the VAR model. The estimated parameter shown is $\mathbf{B}_{1,2}$, which equals zero under the data generating process. The p-values correspond to a test of the null hypothesis that $\mathbf{B}_{1,2} = 0$. Since the null hypothesis is true under the data generating process, valid p-values should follow a uniform distribution.

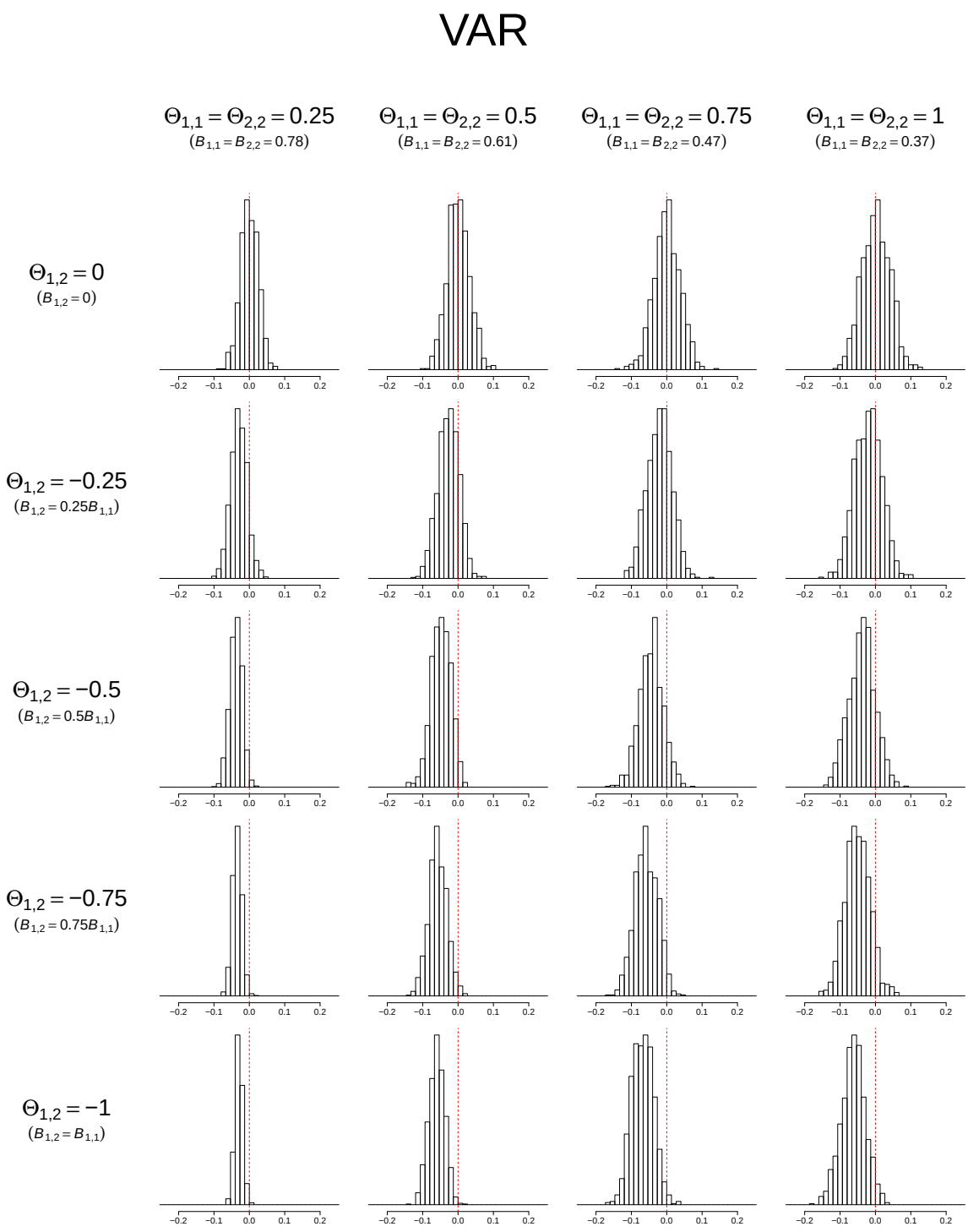
H.3. *Additional graphs for Monte Carlo studies*

Figure A17: Histogram of distribution of estimates from VAR models in temporal aggregation Monte Carlo study with 250 observations



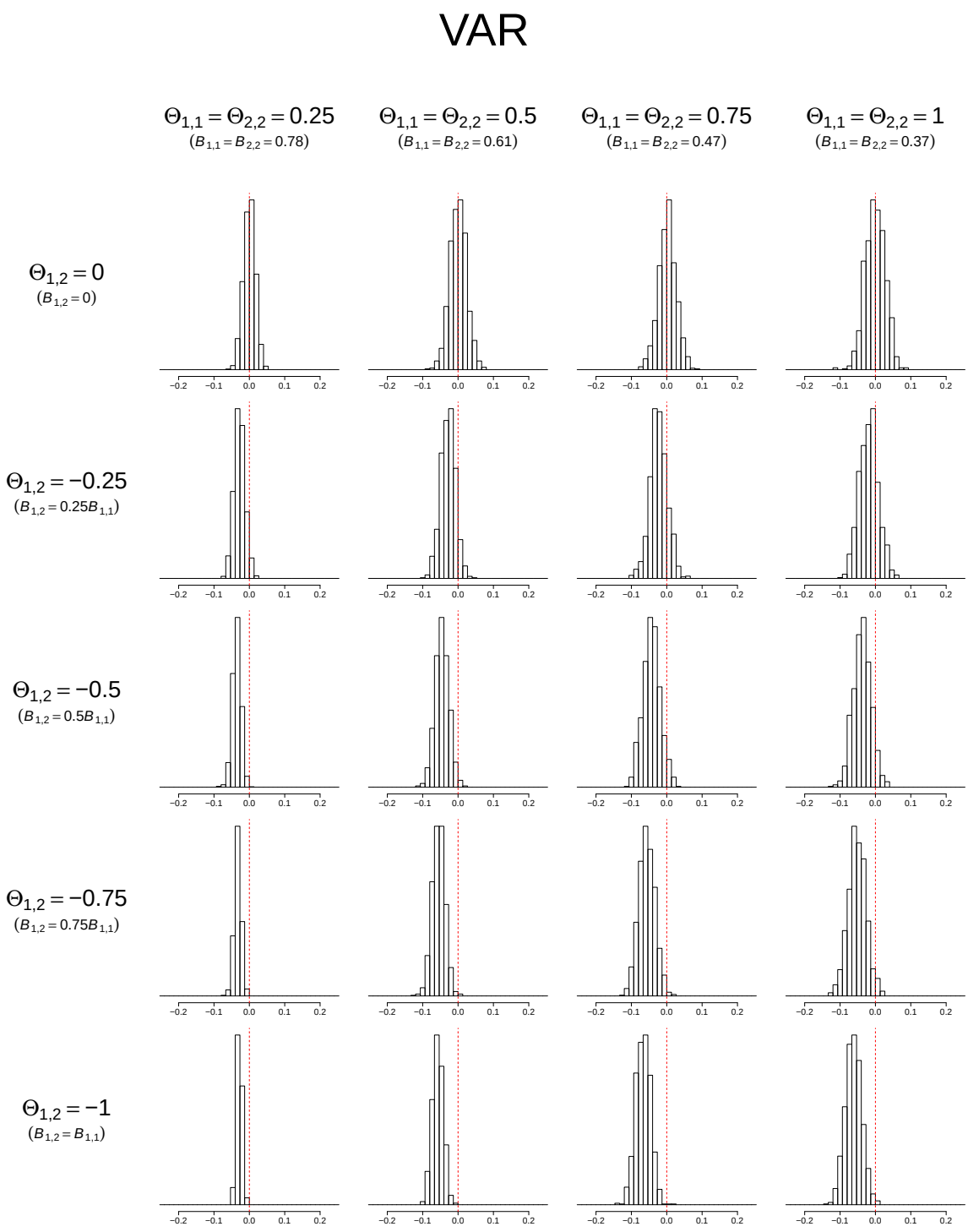
Note: 1,000 simulations per study. In each simulation, the data consist of 250 observations per series. The estimated parameter shown is $\mathbf{B}_{1,2}$, which equals zero under the data generating process. A dashed red line shows this true value of $\mathbf{B}_{1,2} = 0$ in each plot.

Figure A18: Histogram of distribution of estimates from VAR models in temporal aggregation Monte Carlo study with 500 observations



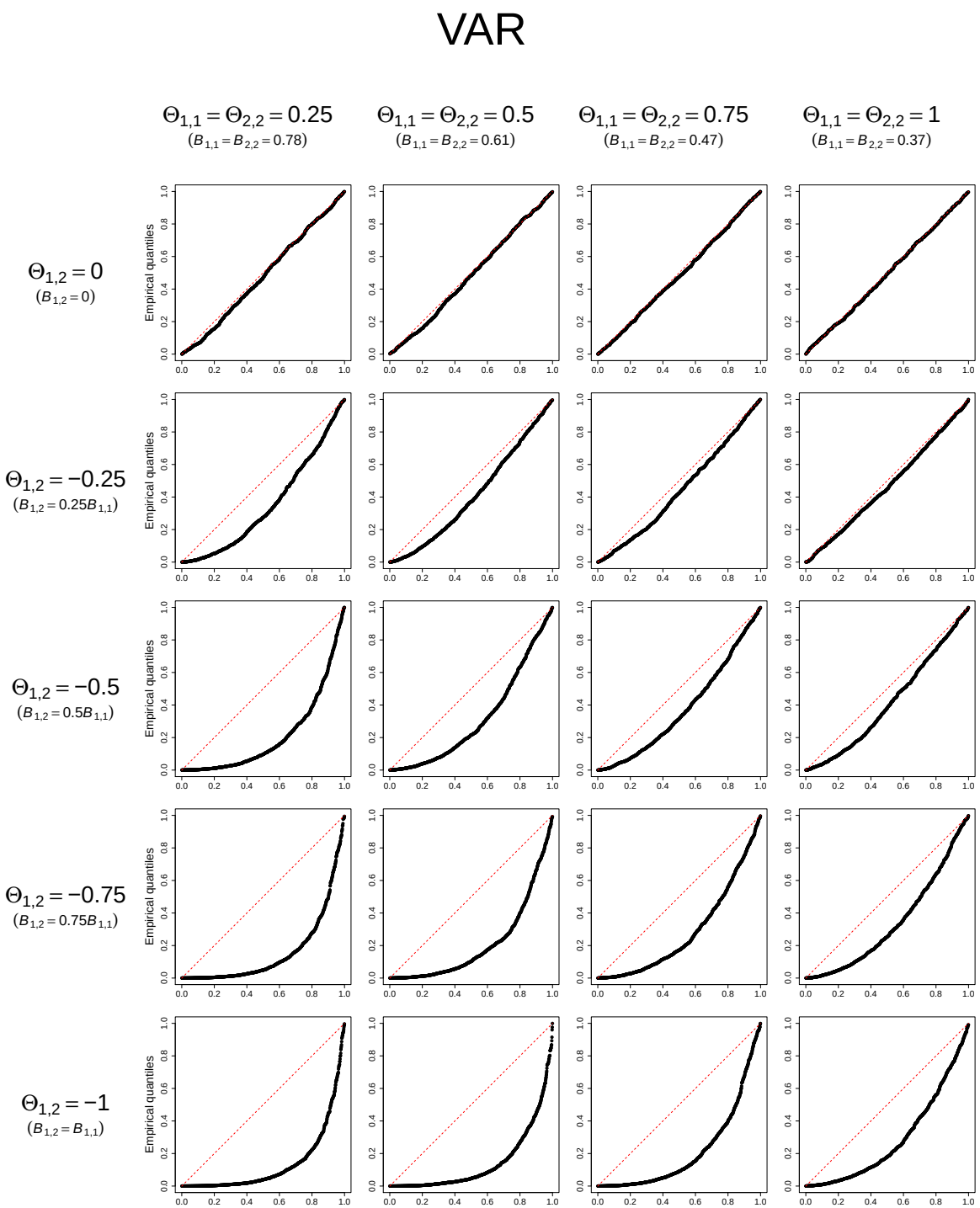
Note: 1,000 simulations per study. In each simulation, the data consist of 500 observations per series. The estimated parameter shown is $\mathbf{B}_{1,2}$, which equals zero under the data generating process. A dashed red line shows this true value of $\mathbf{B}_{1,2} = 0$ in each plot.

Figure A19: Histogram of distribution of estimates from VAR models in temporal aggregation Monte Carlo study with 1,000 observations



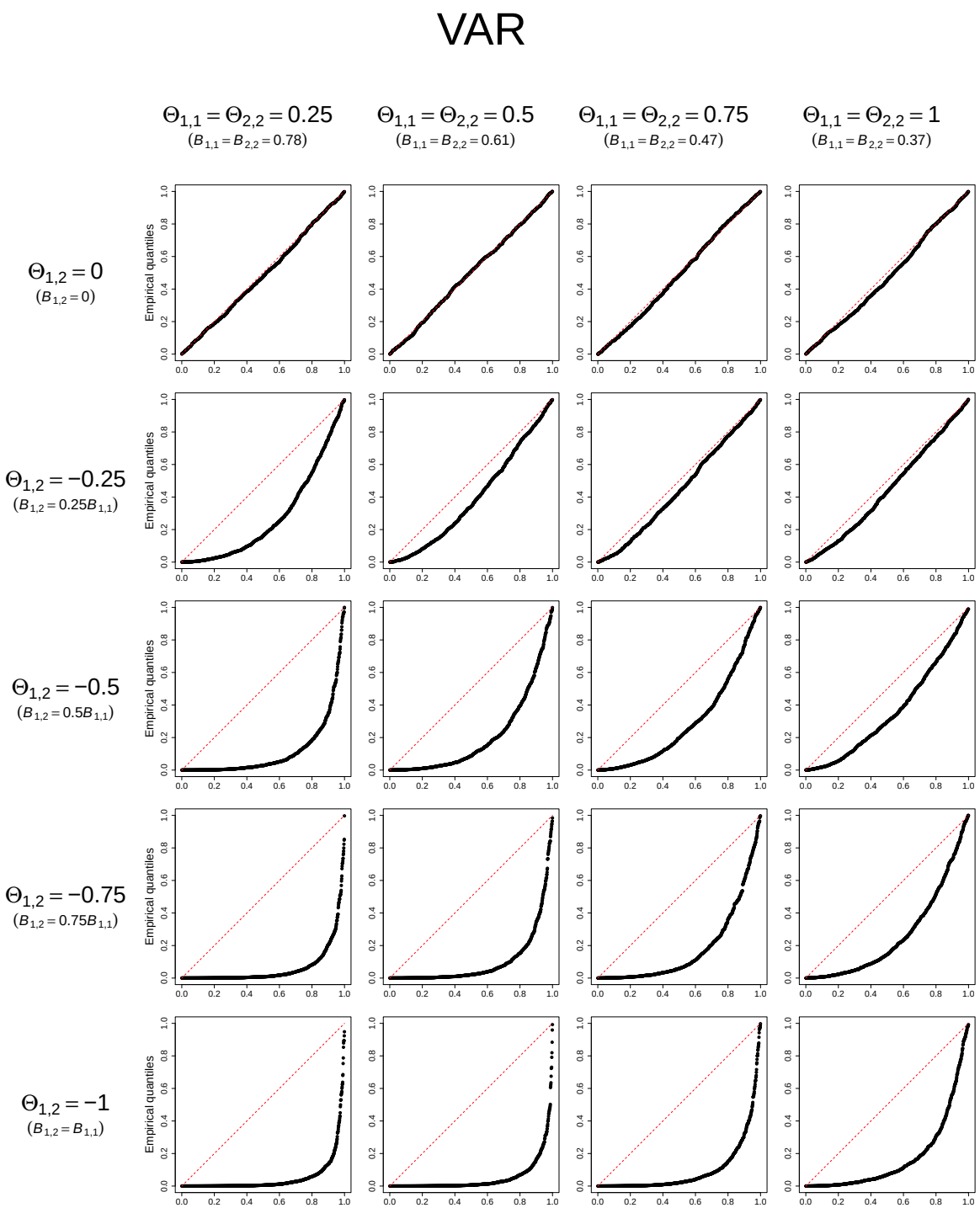
Note: 1,000 simulations per study. In each simulation, the data consist of 1,000 observations per series. The estimated parameter shown is $\mathbf{B}_{1,2}$, which equals zero under the data generating process. A dashed red line shows this true value of $\mathbf{B}_{1,2} = 0$ in each plot.

Figure A20: Quantile-quantile plot of p-values from VAR models in temporal aggregation Monte Carlo study with 250 observations



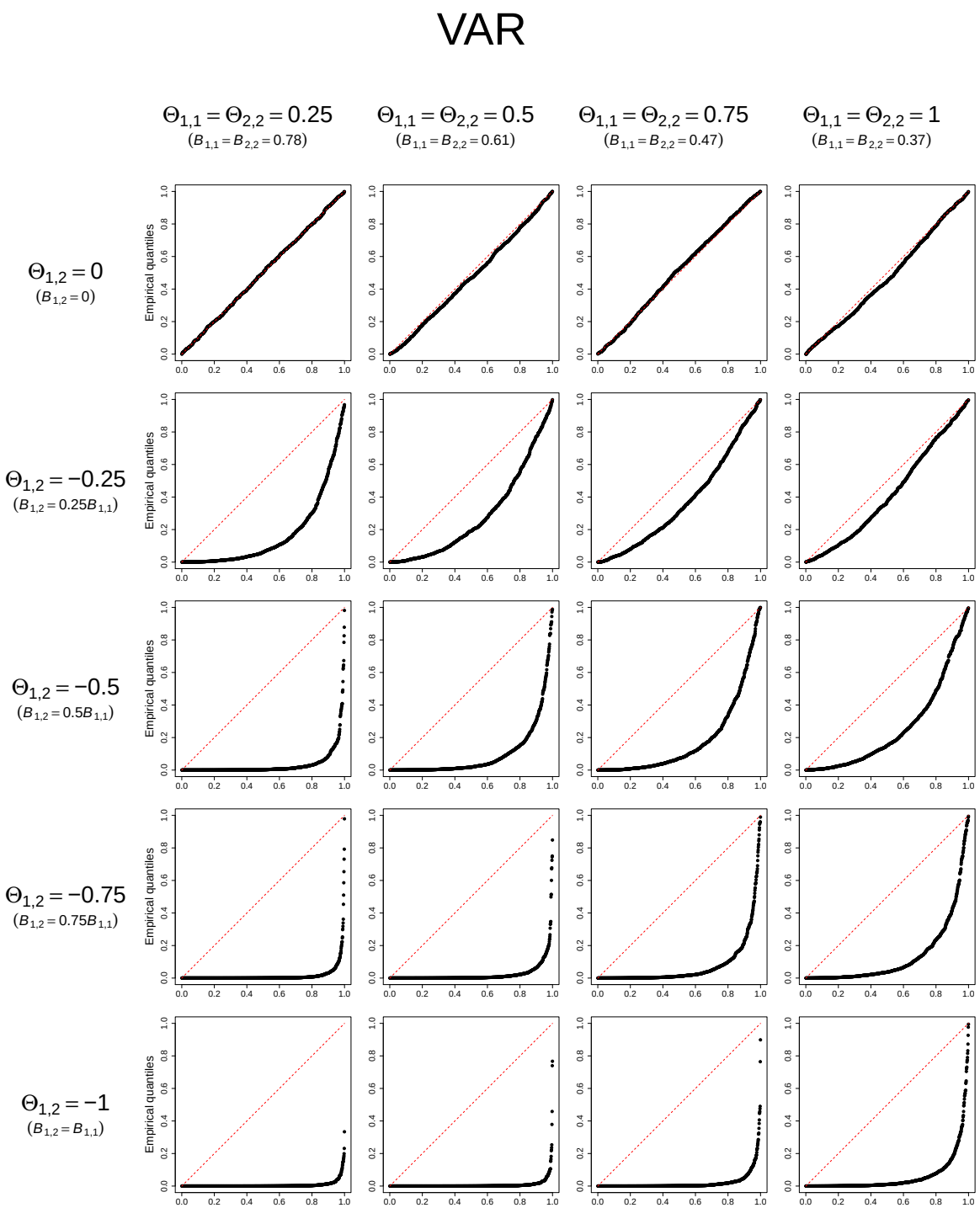
Note: 1,000 simulations per study. In each simulation, the data consist of 250 observations per series. The estimated parameter shown is $\mathbf{B}_{1,2}$, which equals zero under the data generating process. The p-values correspond to a test of the null hypothesis that $\mathbf{B}_{1,2} = 0$. Since the null hypothesis is true under the data generating process, valid p-values should follow a uniform distribution.

Figure A21: Quantile-quantile plot of p-values from VAR models in temporal aggregation Monte Carlo study with 500 observations



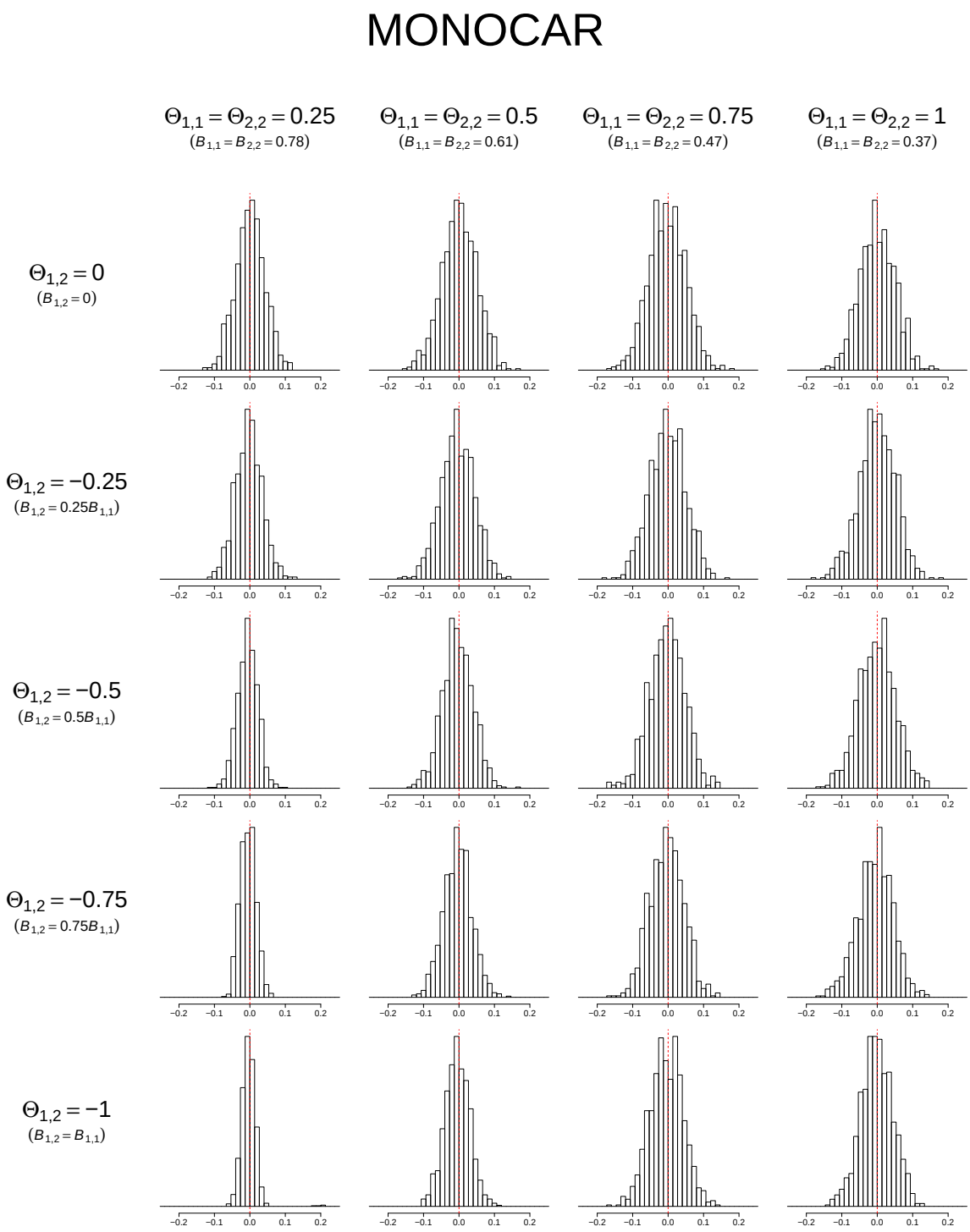
Note: 1,000 simulations per study. In each simulation, the data consist of 500 observations per series. The estimated parameter shown is $\mathbf{B}_{1,2}$, which equals zero under the data generating process. The p-values correspond to a test of the null hypothesis that $\mathbf{B}_{1,2} = 0$. Since the null hypothesis is true under the data generating process, valid p-values should follow a uniform distribution.

Figure A22: Quantile-quantile plot of p-values from VAR models in temporal aggregation Monte Carlo study with 1,000 observations



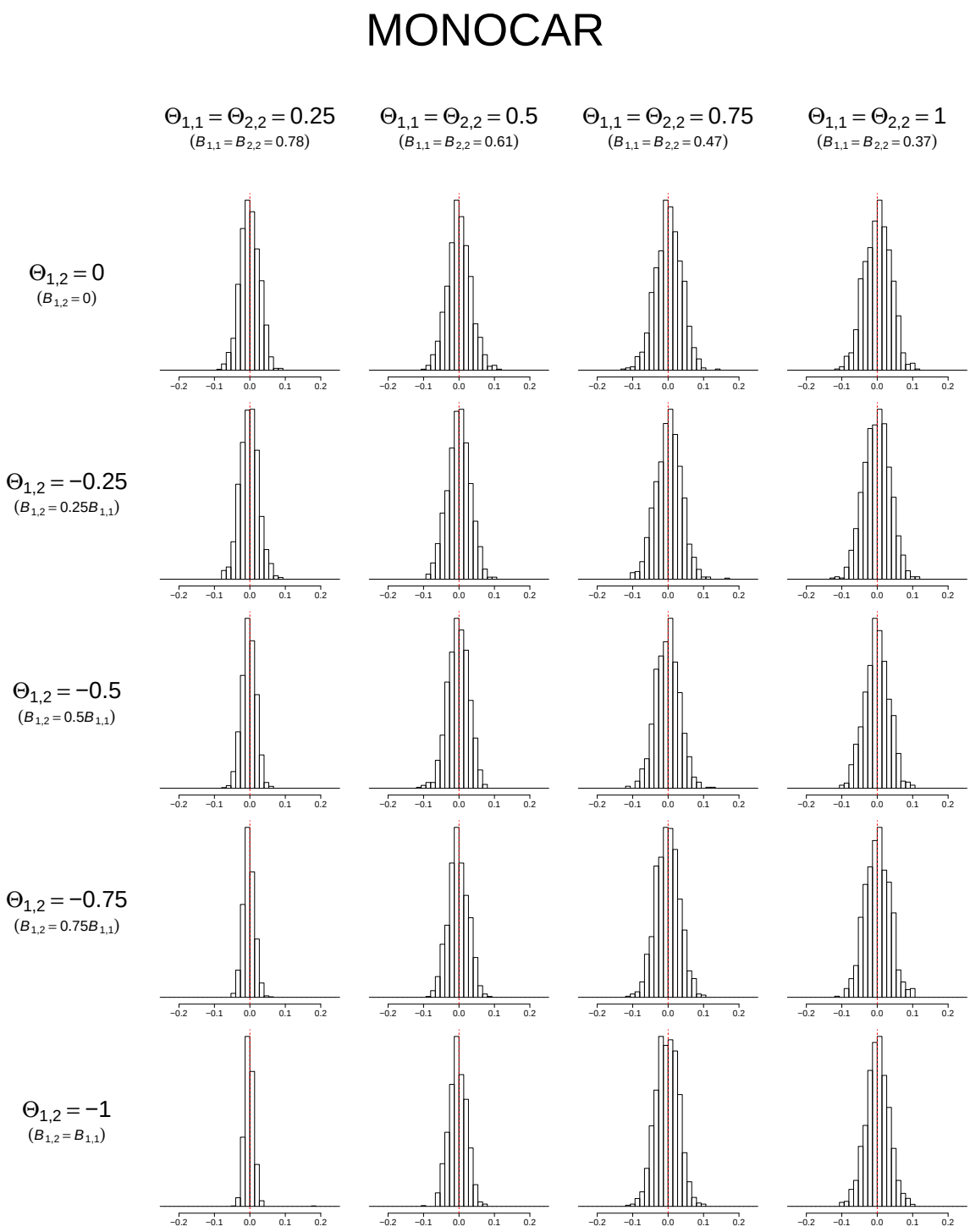
Note: 1,000 simulations per study. In each simulation, the data consist of 1,000 observations per series. The estimated parameter shown is $\mathbf{B}_{1,2}$, which equals zero under the data generating process. The p-values correspond to a test of the null hypothesis that $\mathbf{B}_{1,2} = 0$. Since the null hypothesis is true under the data generating process, valid p-values should follow a uniform distribution.

Figure A23: Histogram of distribution of estimates from MONOCAR models in temporal aggregation Monte Carlo study with 250 observations



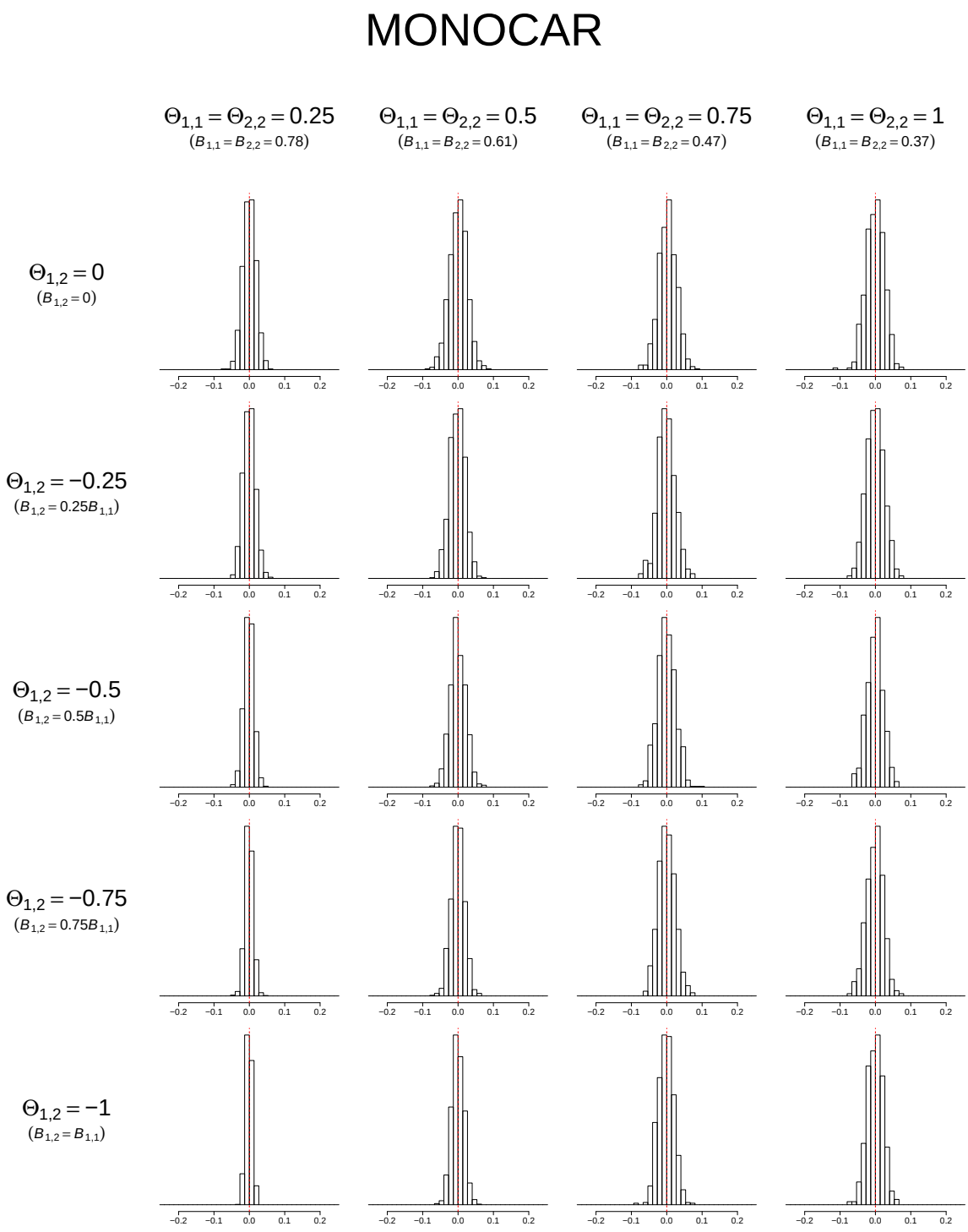
Note: 1,000 simulations per study. In each simulation, the data consist of 250 observations per series. MONOCAR estimates are transformed to use the same parameterization as the VAR model. The estimated parameter shown is $\mathbf{B}_{1,2}$, which equals zero under the data generating process. A dashed red line shows this true value of $\mathbf{B}_{1,2} = 0$ in each plot.

Figure A24: Histogram of distribution of estimates from MONOCAR models in temporal aggregation Monte Carlo study with 500 observations



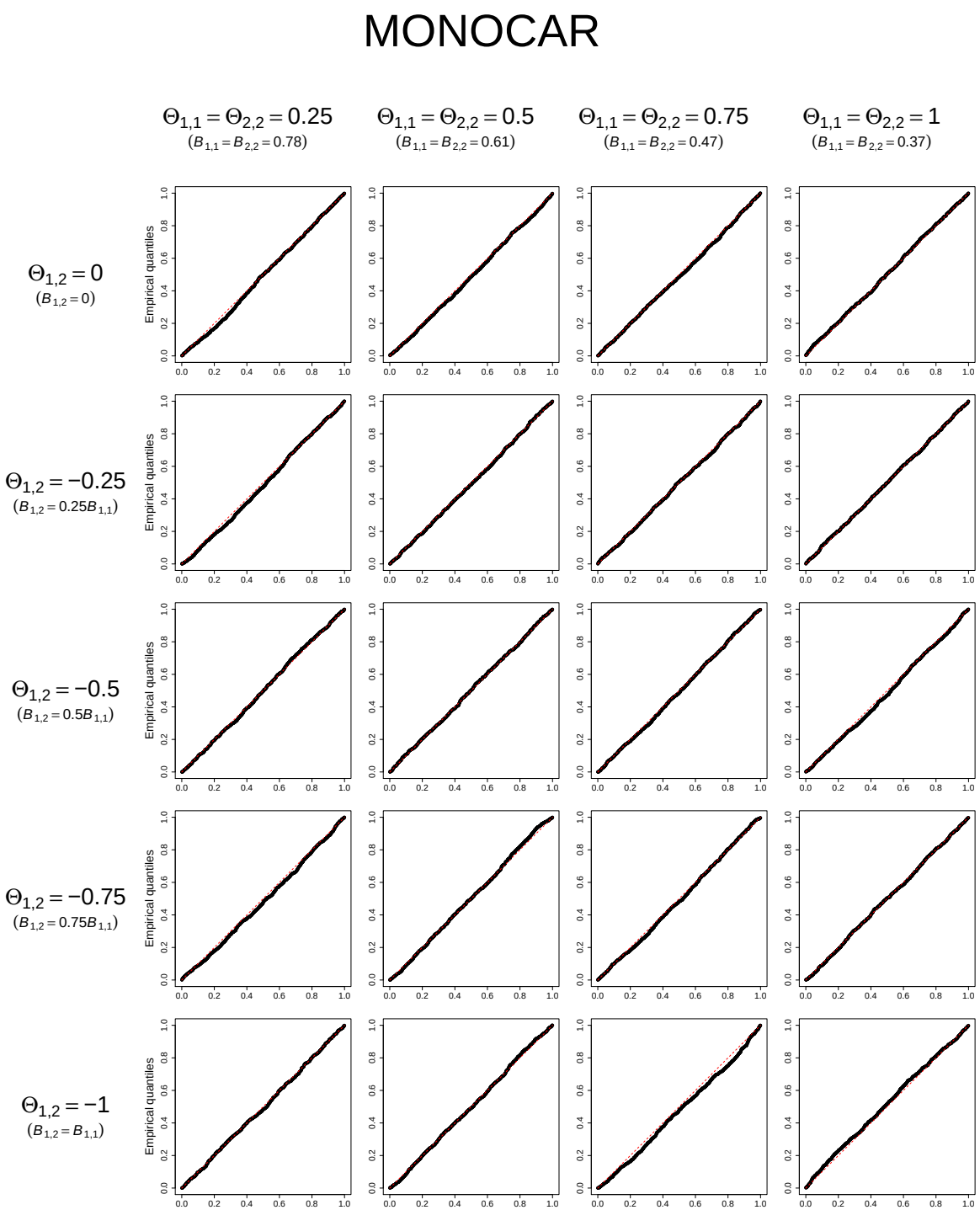
Note: 1,000 simulations per study. In each simulation, the data consist of 500 observations per series. MONOCAR estimates are transformed to use the same parameterization as the VAR model. The estimated parameter shown is $\mathbf{B}_{1,2}$, which equals zero under the data generating process. A dashed red line shows this true value of $\mathbf{B}_{1,2} = 0$ in each plot.

Figure A25: Histogram of distribution of estimates from MONOCAR models in temporal aggregation Monte Carlo study with 1,000 observations



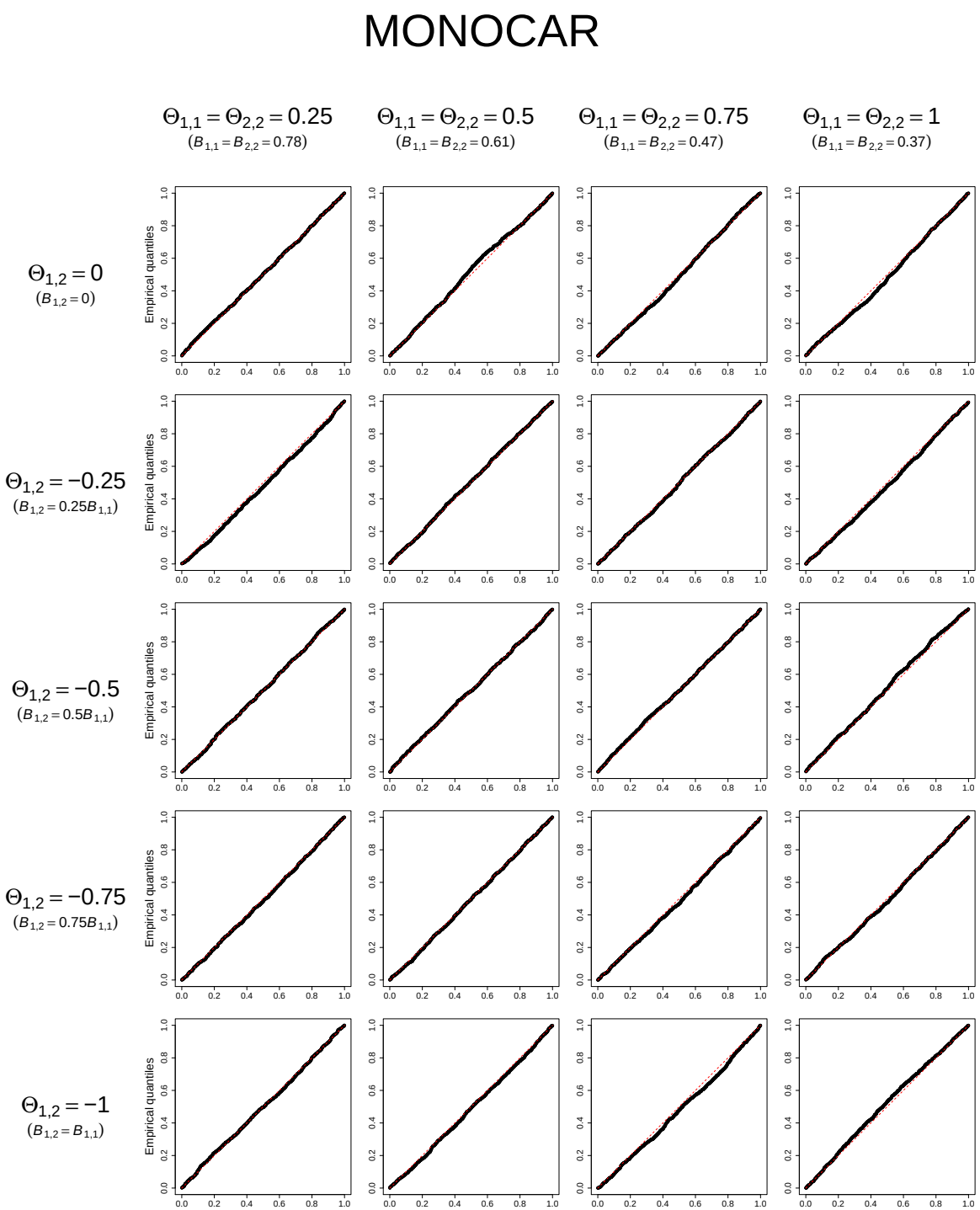
Note: 1,000 simulations per study. In each simulation, the data consist of 1,000 observations per series. MONOCAR estimates are transformed to use the same parameterization as the VAR model. The estimated parameter shown is $\mathbf{B}_{1,2}$, which equals zero under the data generating process. A dashed red line shows this true value of $\mathbf{B}_{1,2} = 0$ in each plot.

Figure A26: Quantile-quantile plot of p-values from MONOCAR models in temporal aggregation Monte Carlo study with 250 observations



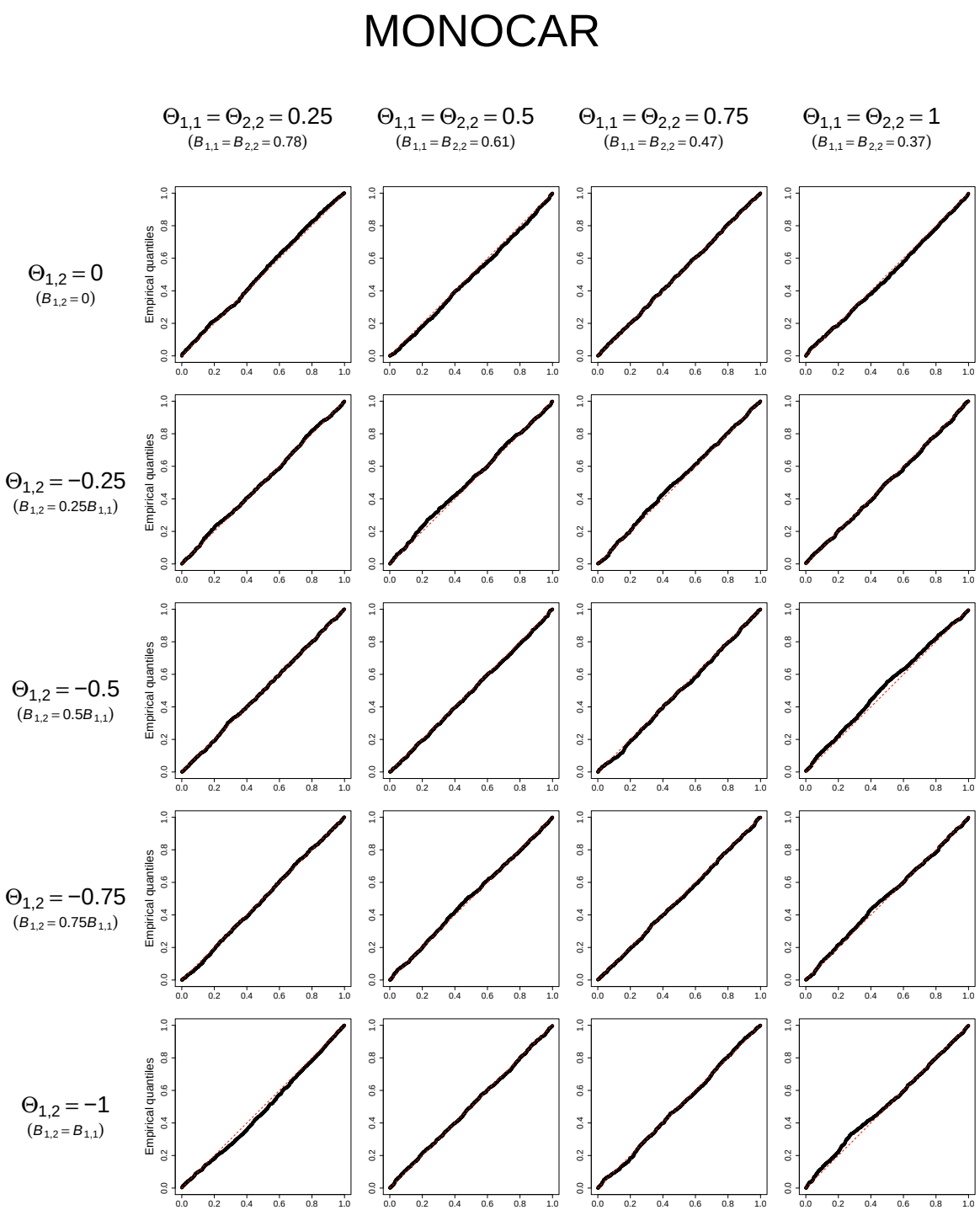
Note: 1,000 simulations per study. In each simulation, the data consist of 250 observations per series. The estimated parameter shown is $\mathbf{B}_{1,2}$, which equals zero under the data generating process. The p-values correspond to a test of the null hypothesis that $\mathbf{B}_{1,2} = 0$. Since the null hypothesis is true under the data generating process, valid p-values should follow a uniform distribution.

Figure A27: Quantile-quantile plot of p-values from MONOCAR models in temporal aggregation Monte Carlo study with 500 observations



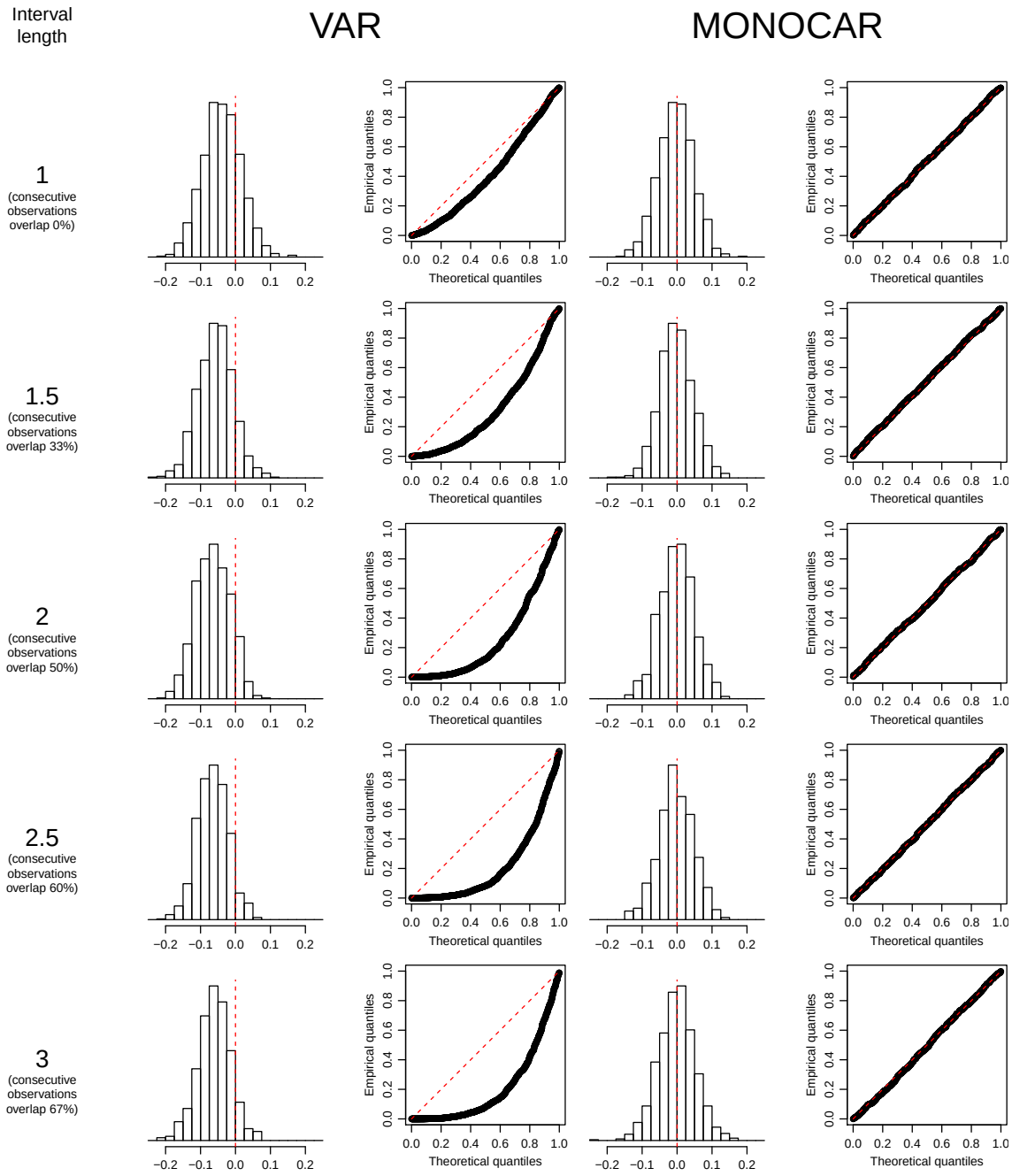
Note: 1,000 simulations per study. In each simulation, the data consist of 500 observations per series. The estimated parameter shown is $\mathbf{B}_{1,2}$, which equals zero under the data generating process. The p-values correspond to a test of the null hypothesis that $\mathbf{B}_{1,2} = 0$. Since the null hypothesis is true under the data generating process, valid p-values should follow a uniform distribution.

Figure A28: Quantile-quantile plot of p-values from MONOCAR models in temporal aggregation Monte Carlo study with 1,000 observations



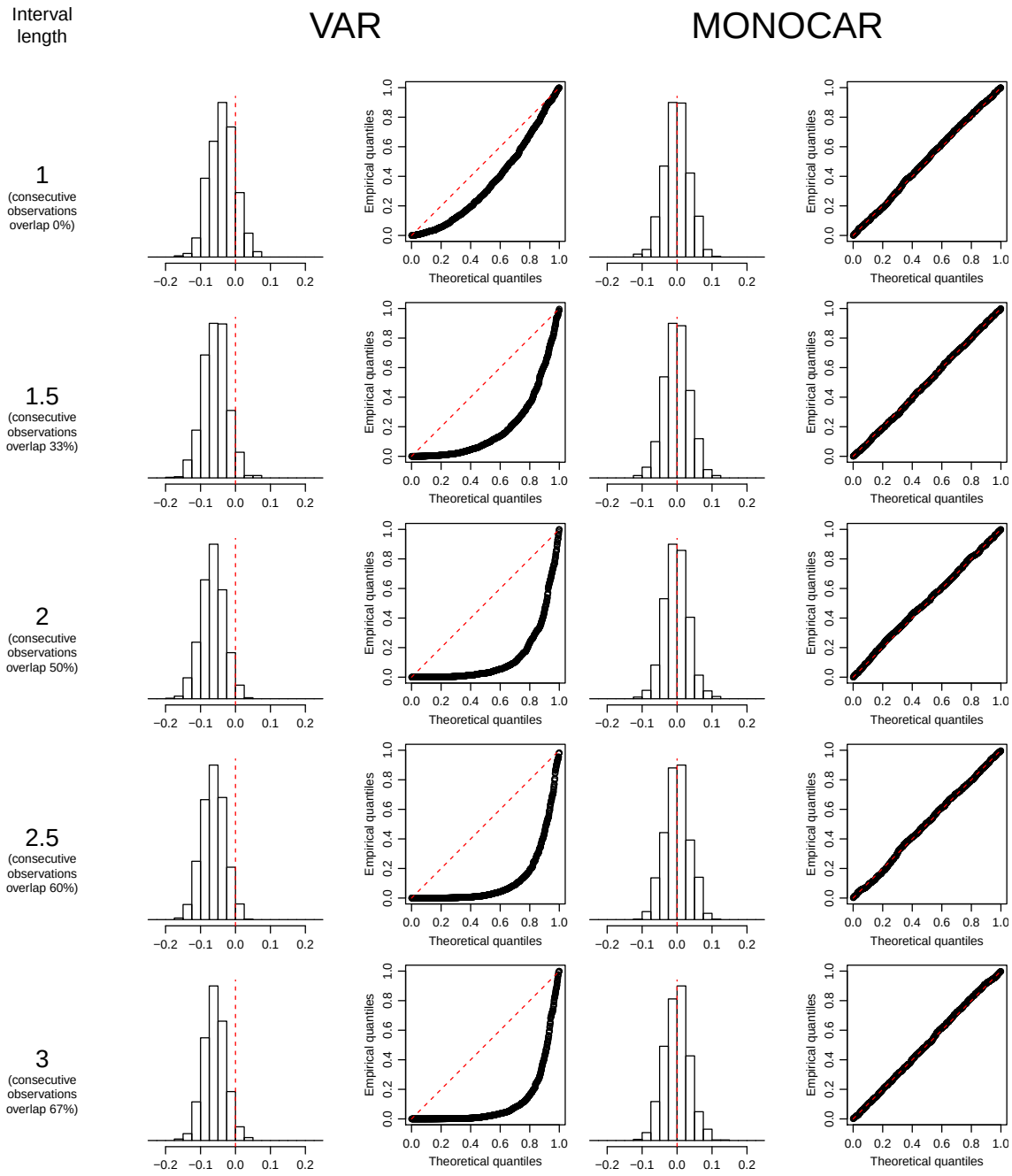
Note: 1,000 simulations per study. In each simulation, the data consist of 1,000 observations per series. The estimated parameter shown is $\mathbf{B}_{1,2}$, which equals zero under the data generating process. The p-values correspond to a test of the null hypothesis that $\mathbf{B}_{1,2} = 0$. Since the null hypothesis is true under the data generating process, valid p-values should follow a uniform distribution.

Figure A29: Histogram of distribution of estimates and quantile-quantile plot of p-values from VAR and MONOCAR models for the overlapping observation Monte Carlo study with 250 observations



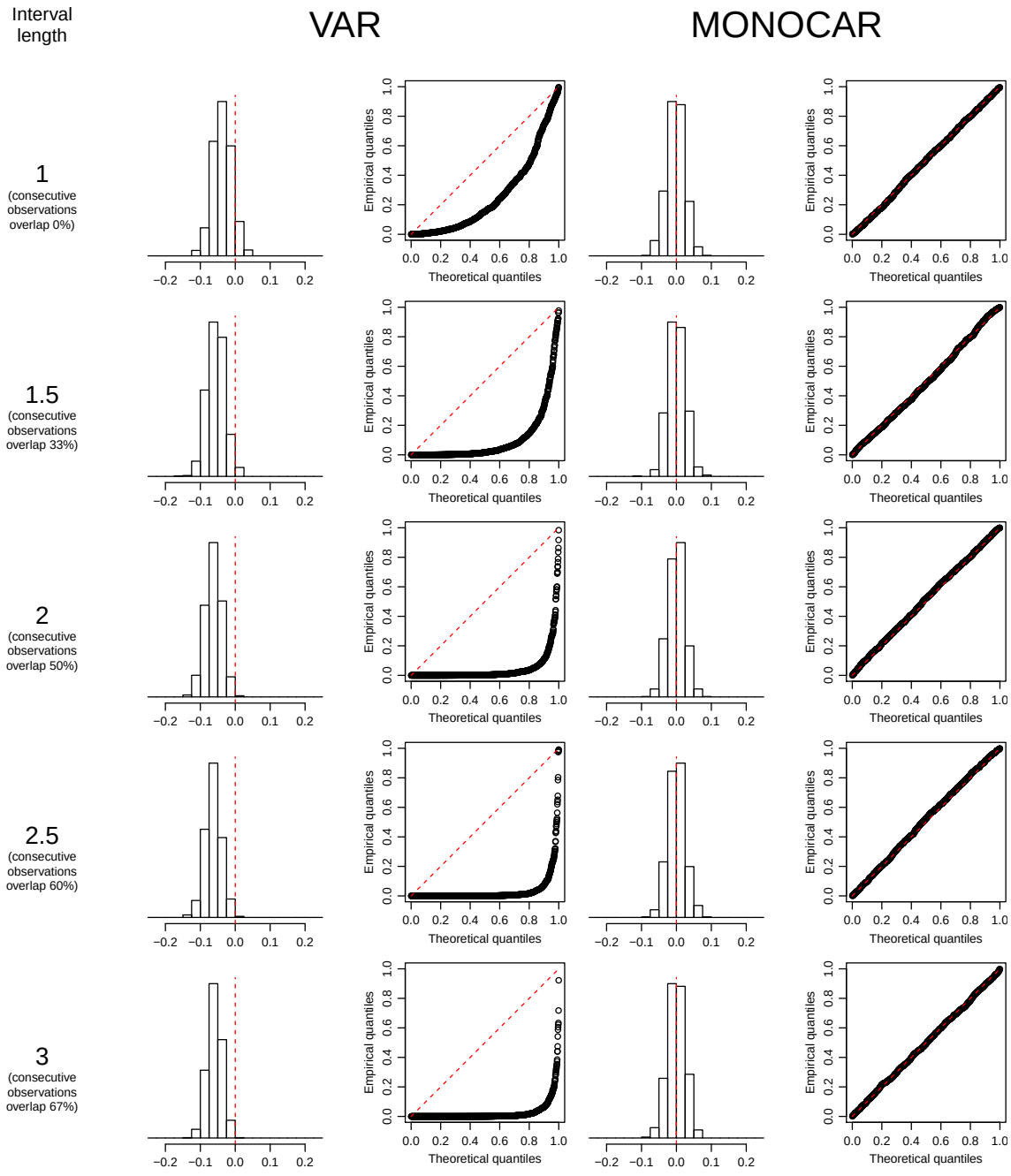
Note: 1,000 simulations per study. In each simulation, the data consist of 250 observations per series. The estimated parameter shown is $\mathbf{B}_{1,2}$, which equals zero under the data generating process. The p-values correspond to a test of the null hypothesis that $\mathbf{B}_{1,2} = 0$. Since the null hypothesis is true under the data generating process, valid p-values should follow a uniform distribution.

Figure A30: Histogram of distribution of estimates and quantile-quantile plot of p-values from VAR and MONOCAR models for the overlapping observation Monte Carlo study with 500 observations



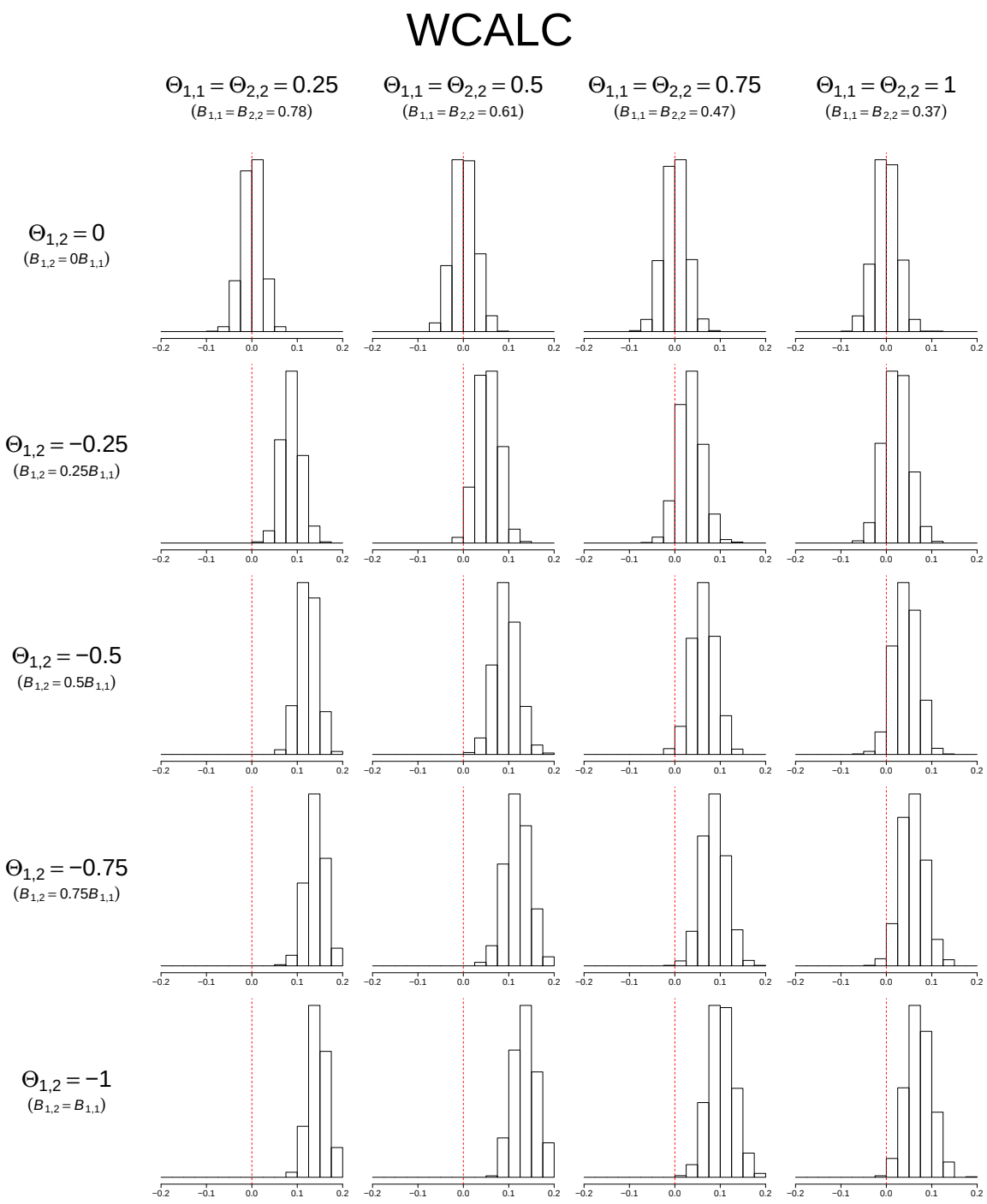
Note: 1,000 simulations per study. In each simulation, the data consist of 500 observations per series. The estimated parameter shown is $\mathbf{B}_{1,2}$, which equals zero under the data generating process. The p-values correspond to a test of the null hypothesis that $\mathbf{B}_{1,2} = 0$. Since the null hypothesis is true under the data generating process, valid p-values should follow a uniform distribution.

Figure A31: Histogram of distribution of estimates and quantile-quantile plot of p-values from VAR and MONOCAR models for the overlapping observation Monte Carlo study with 1,000 observations



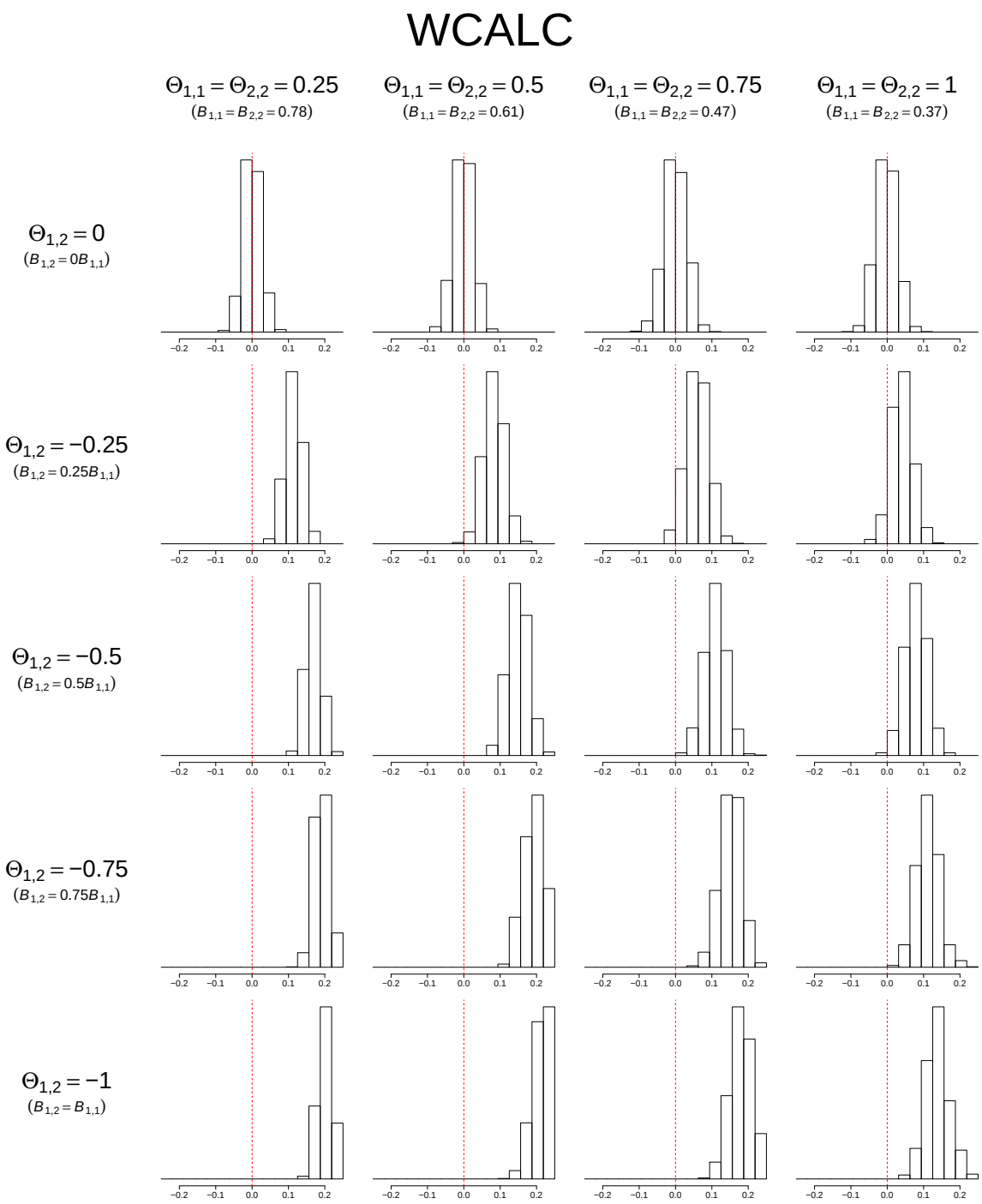
Note: 1,000 simulations per study. In each simulation, the data consist of 1,000 observations per series. The estimated parameter shown is $\mathbf{B}_{1,2}$, which equals zero under the data generating process. The p-values correspond to a test of the null hypothesis that $\mathbf{B}_{1,2} = 0$. Since the null hypothesis is true under the data generating process, valid p-values should follow a uniform distribution.

Figure A32: Histogram of distribution of estimates from WCALC and VAR models in imputation Monte Carlo study with average of four polls per year



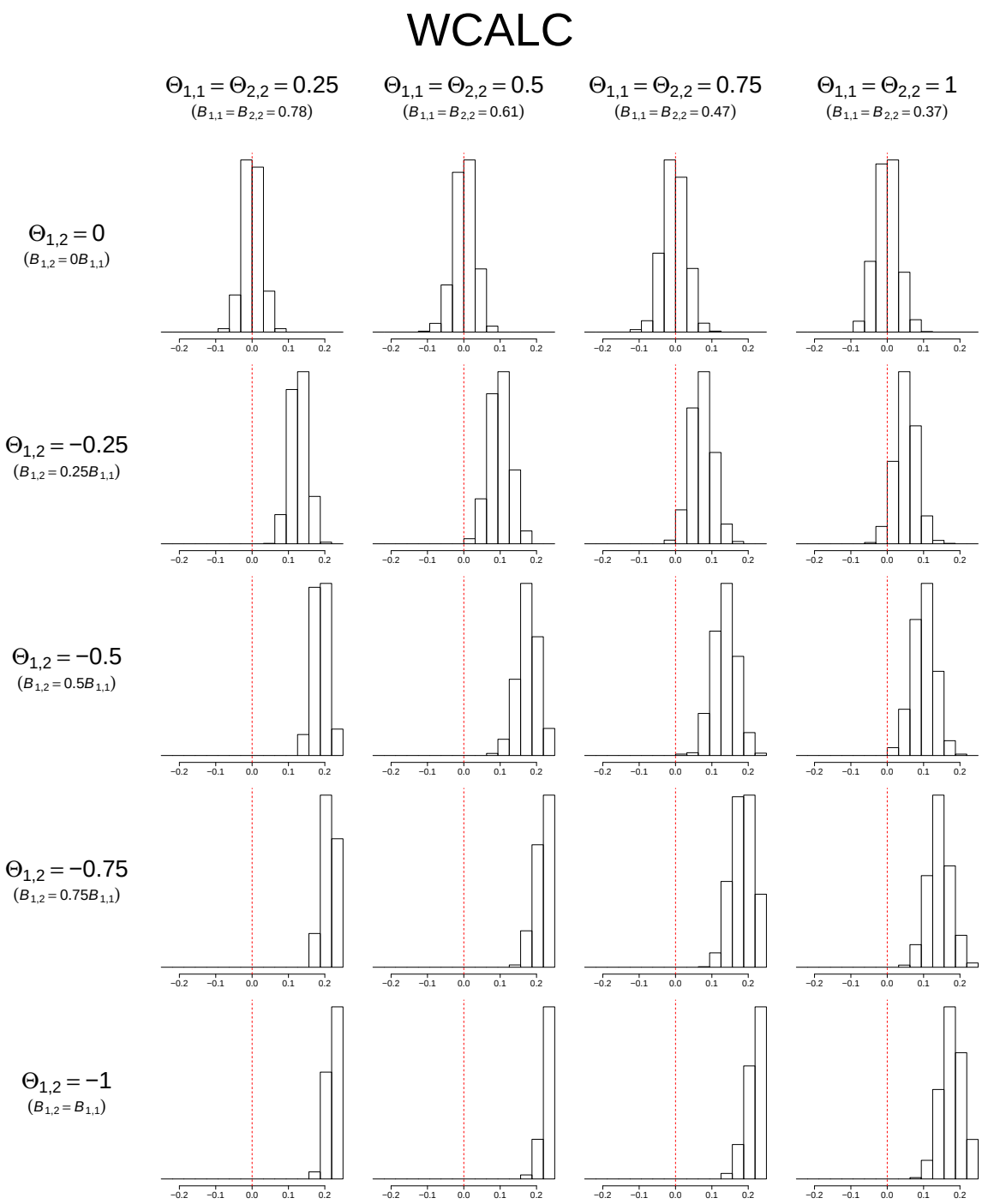
Note: 1,000 simulations per study. In each simulation, the first series consists of 160 randomly spaced instantaneous polls (averaging four per year) and the second series consists of monthly instantaneous observations. The estimated parameter shown is $\mathbf{B}_{1,2}$, which equals zero under the data generating process. A dashed red line shows this true value of $\mathbf{B}_{1,2} = 0$ in each plot.

Figure A33: Histogram of distribution of estimates from WCALC and VAR models in imputation Monte Carlo study with average of eight polls per year



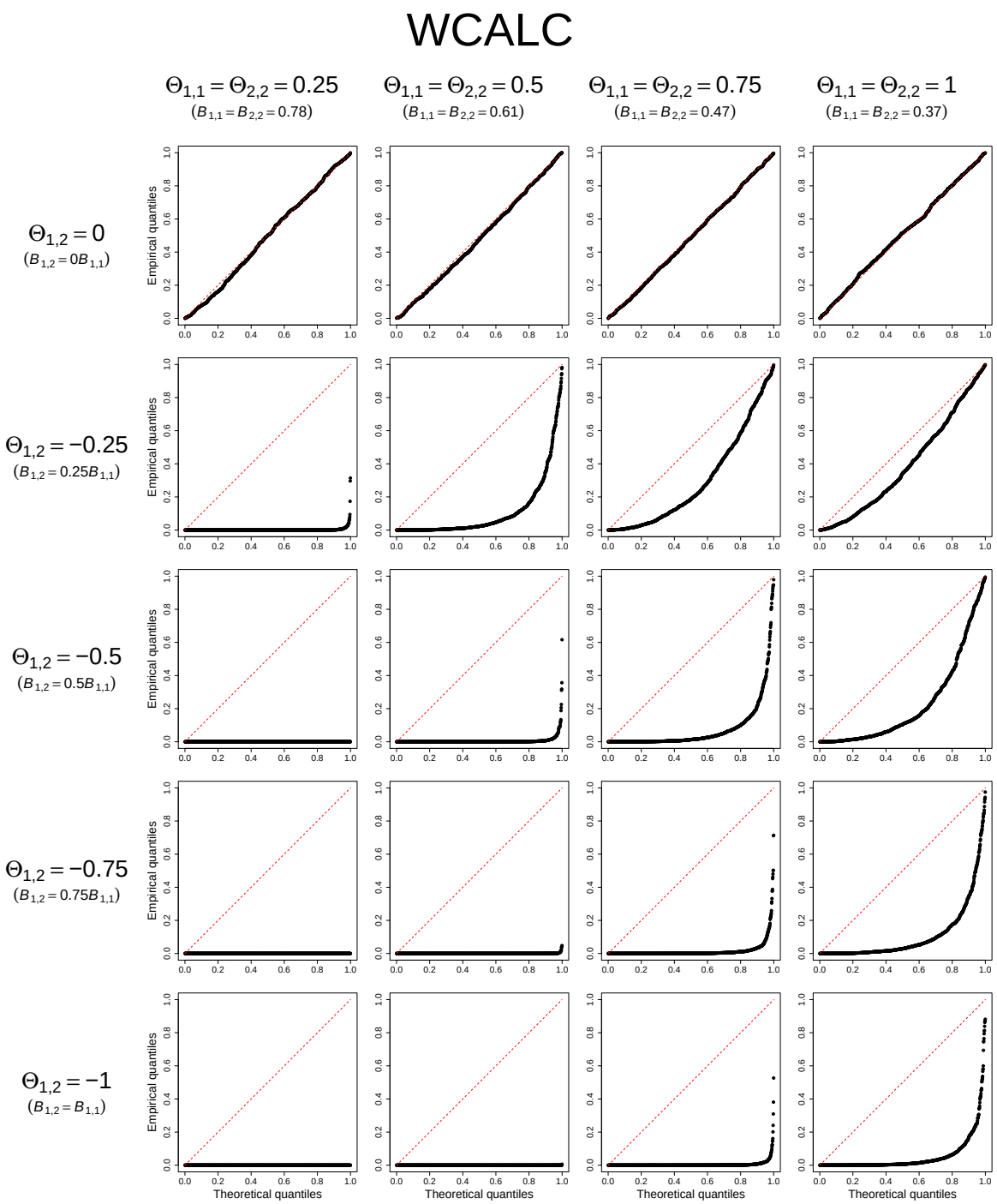
Note: 1,000 simulations per study. In each simulation, the first series consists of 320 randomly spaced instantaneous polls (averaging eight per year) and the second series consists of monthly instantaneous observations. The estimated parameter shown is $\mathbf{B}_{1,2}$, which equals zero under the data generating process. A dashed red line shows this true value of $\mathbf{B}_{1,2} = 0$ in each plot.

Figure A34: Histogram of distribution of estimates from WCALC and VAR models in imputation Monte Carlo study with average of twelve polls per year



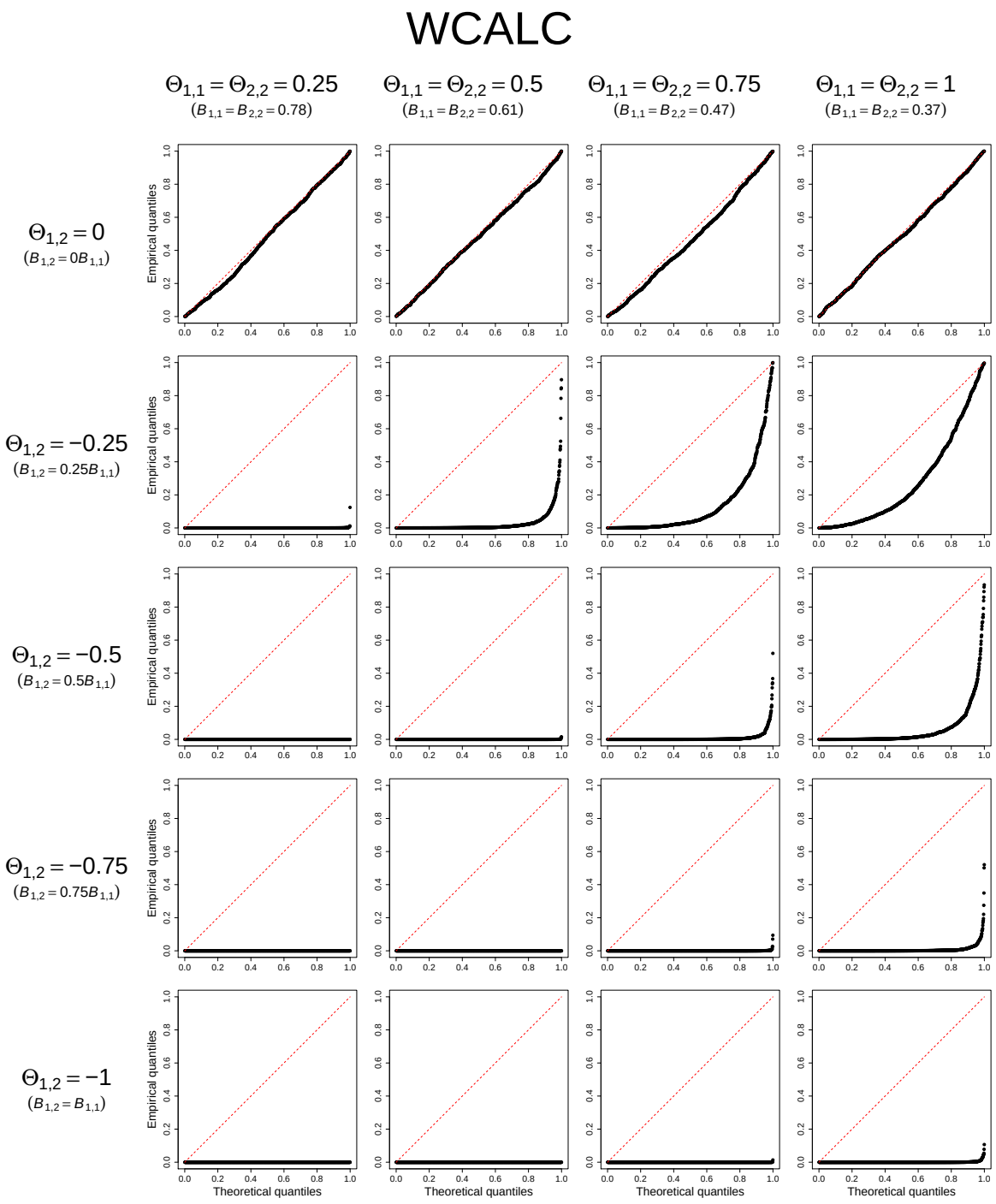
Note: 1,000 simulations per study. In each simulation, the first series consists of 480 randomly spaced instantaneous polls (averaging twelve per year) and the second series consists of monthly instantaneous observations. The estimated parameter shown is $\mathbf{B}_{1,2}$, which equals zero under the data generating process. A dashed red line shows this true value of $\mathbf{B}_{1,2} = 0$ in each plot.

Figure A35: Quantile-quantile plot of p-values from WCALC and VAR models in imputation Monte Carlo study with average of four polls per year



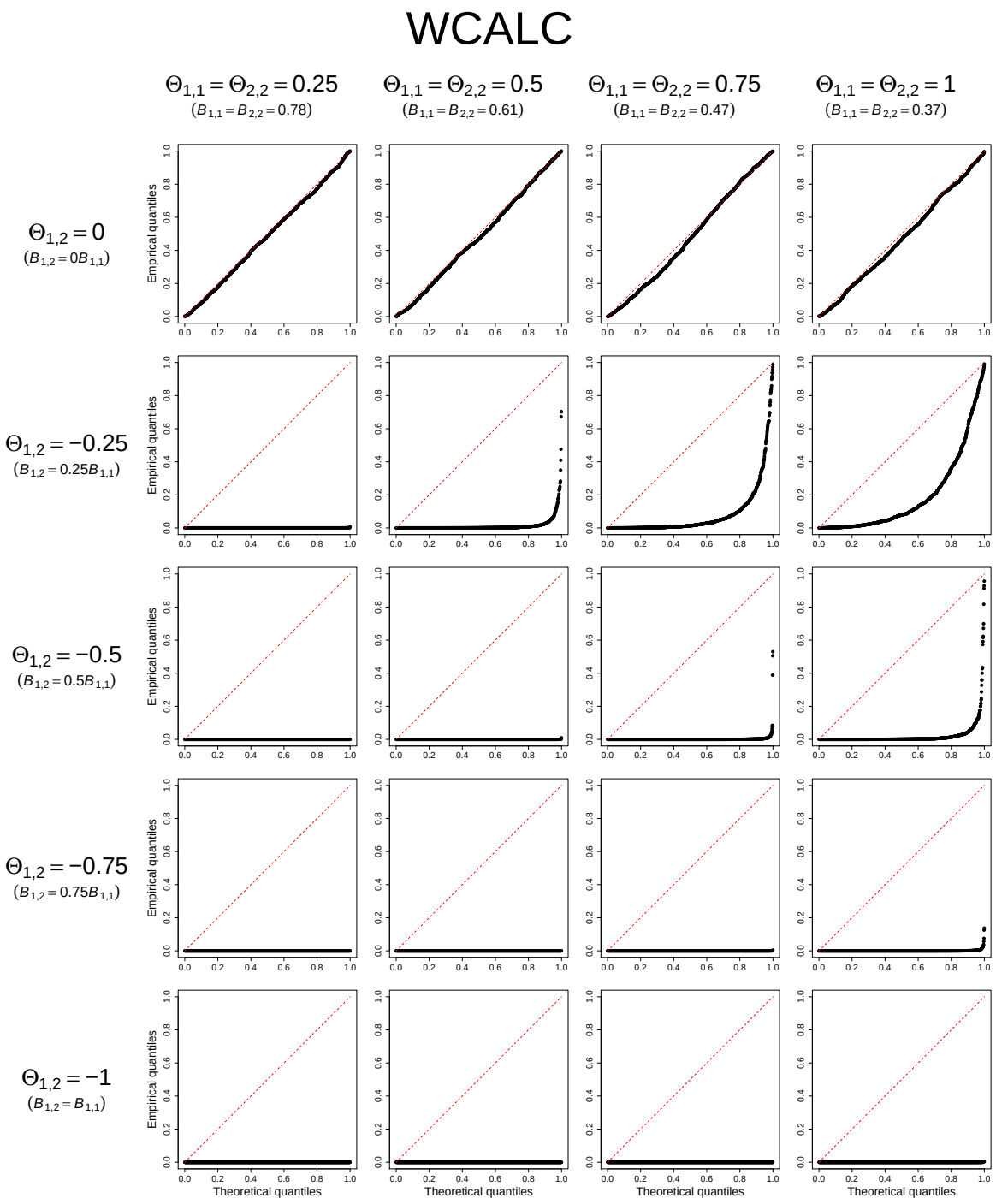
Note: 1,000 simulations per study. In each simulation, the first series consists of 160 randomly spaced instantaneous polls (averaging four per year) and the second series consists of monthly instantaneous observations. The estimated parameter shown is $\mathbf{B}_{1,2}$, which equals zero under the data generating process. The p-values correspond to a test of the null hypothesis that $\mathbf{B}_{1,2} = 0$. Since the null hypothesis is true under the data generating process, valid p-values should follow a uniform distribution.

Figure A36: Quantile-quantile plot of p-values from WCALC and VAR models in imputation Monte Carlo study with average of eight polls per year



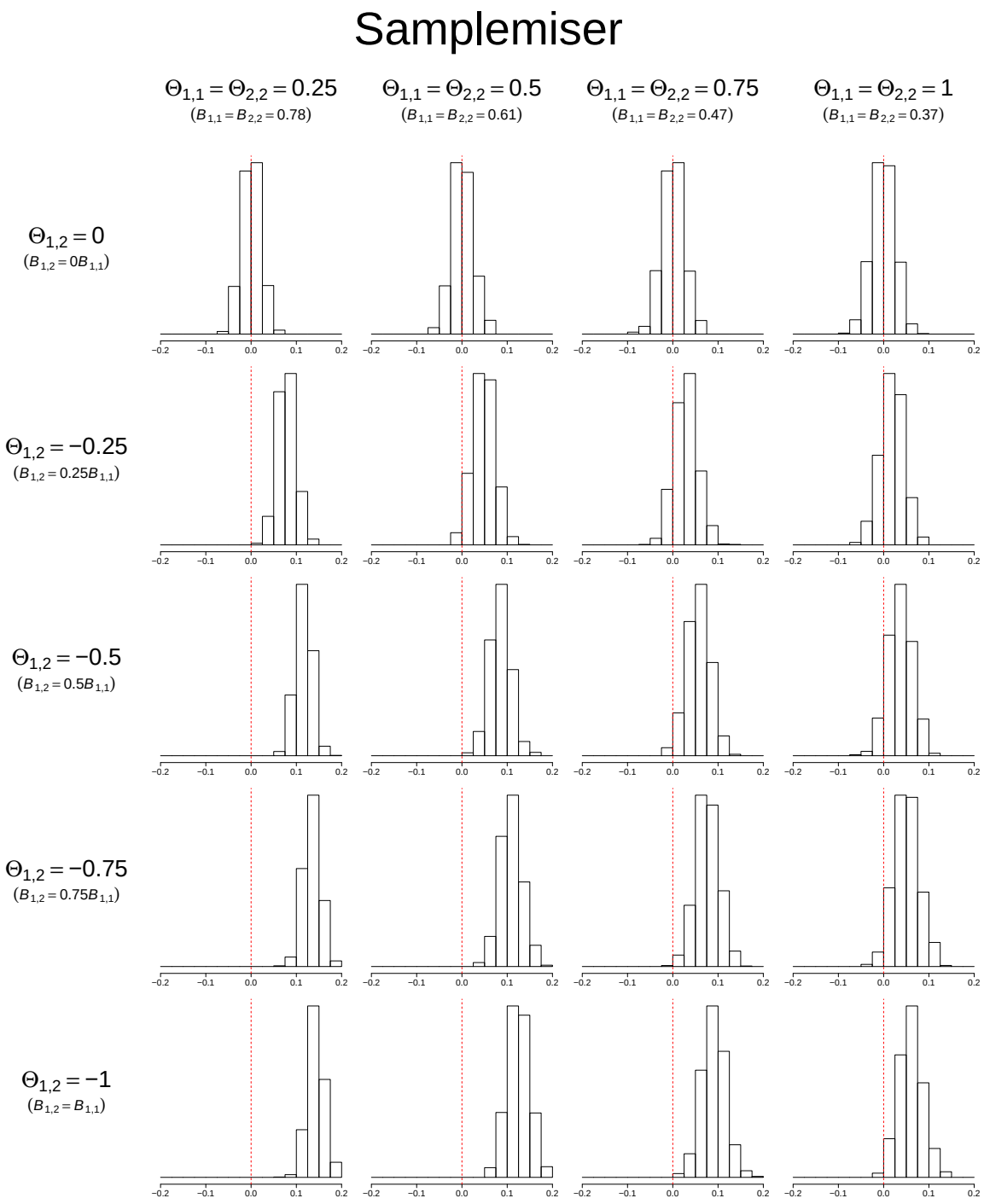
Note: 1,000 simulations per study. In each simulation, the first series consists of 320 randomly spaced instantaneous polls (averaging eight per year) and the second series consists of monthly instantaneous observations. The estimated parameter shown is $\mathbf{B}_{1,2}$, which equals zero under the data generating process. The p-values correspond to a test of the null hypothesis that $\mathbf{B}_{1,2} = 0$. Since the null hypothesis is true under the data generating process, valid p-values should follow a uniform distribution.

Figure A37: Quantile-quantile plot of p-values from WCALC and VAR models in imputation Monte Carlo study with average of twelve polls per year



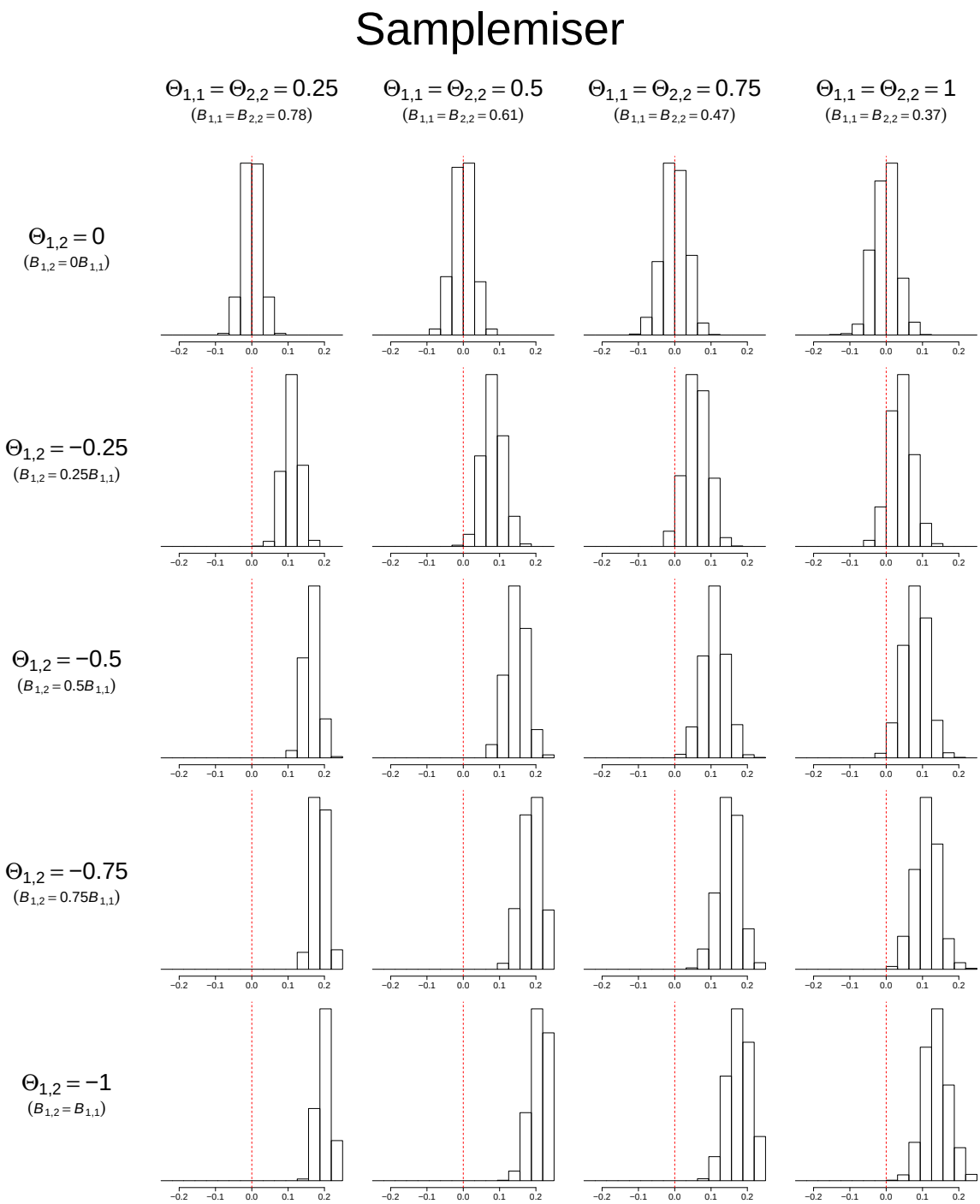
Note: 1,000 simulations per study. In each simulation, the first series consists of 480 randomly spaced instantaneous polls (averaging twelve per year) and the second series consists of monthly instantaneous observations. The estimated parameter shown is $\mathbf{B}_{1,2}$, which equals zero under the data generating process. The p-values correspond to a test of the null hypothesis that $\mathbf{B}_{1,2} = 0$. Since the null hypothesis is true under the data generating process, valid p-values should follow a uniform distribution.

Figure A38: Histogram of distribution of estimates from Samplemiser and VAR models in imputation Monte Carlo study with average of four polls per year



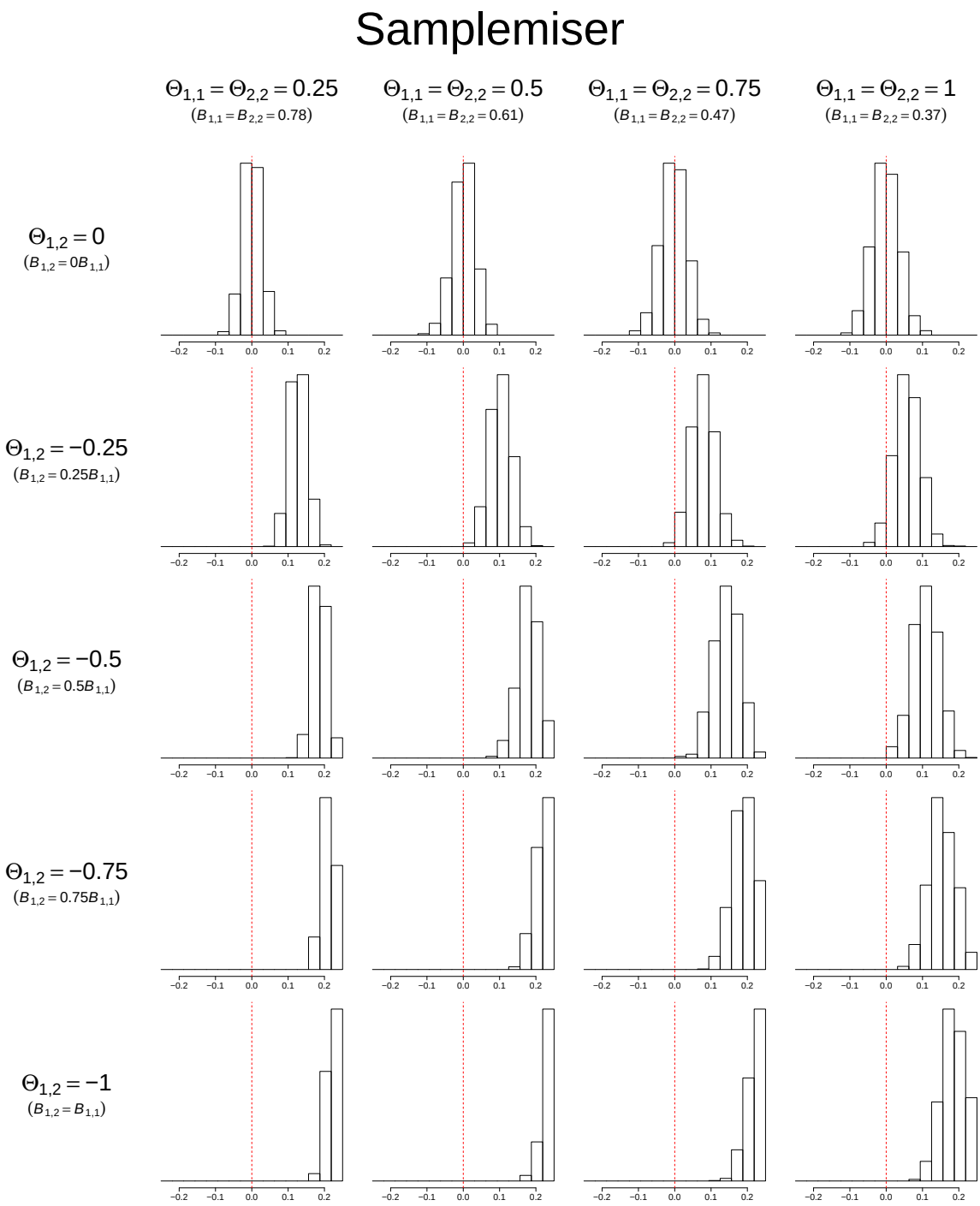
Note: 1,000 simulations per study. In each simulation, the first series consists of 160 randomly spaced instantaneous polls (averaging four per year) and the second series consists of monthly instantaneous observations. The estimated parameter shown is $\mathbf{B}_{1,2}$, which equals zero under the data generating process. A dashed red line shows this true value of $\mathbf{B}_{1,2} = 0$ in each plot.

Figure A39: Histogram of distribution of estimates from Samplemiser and VAR models in imputation Monte Carlo study with average of eight polls per year



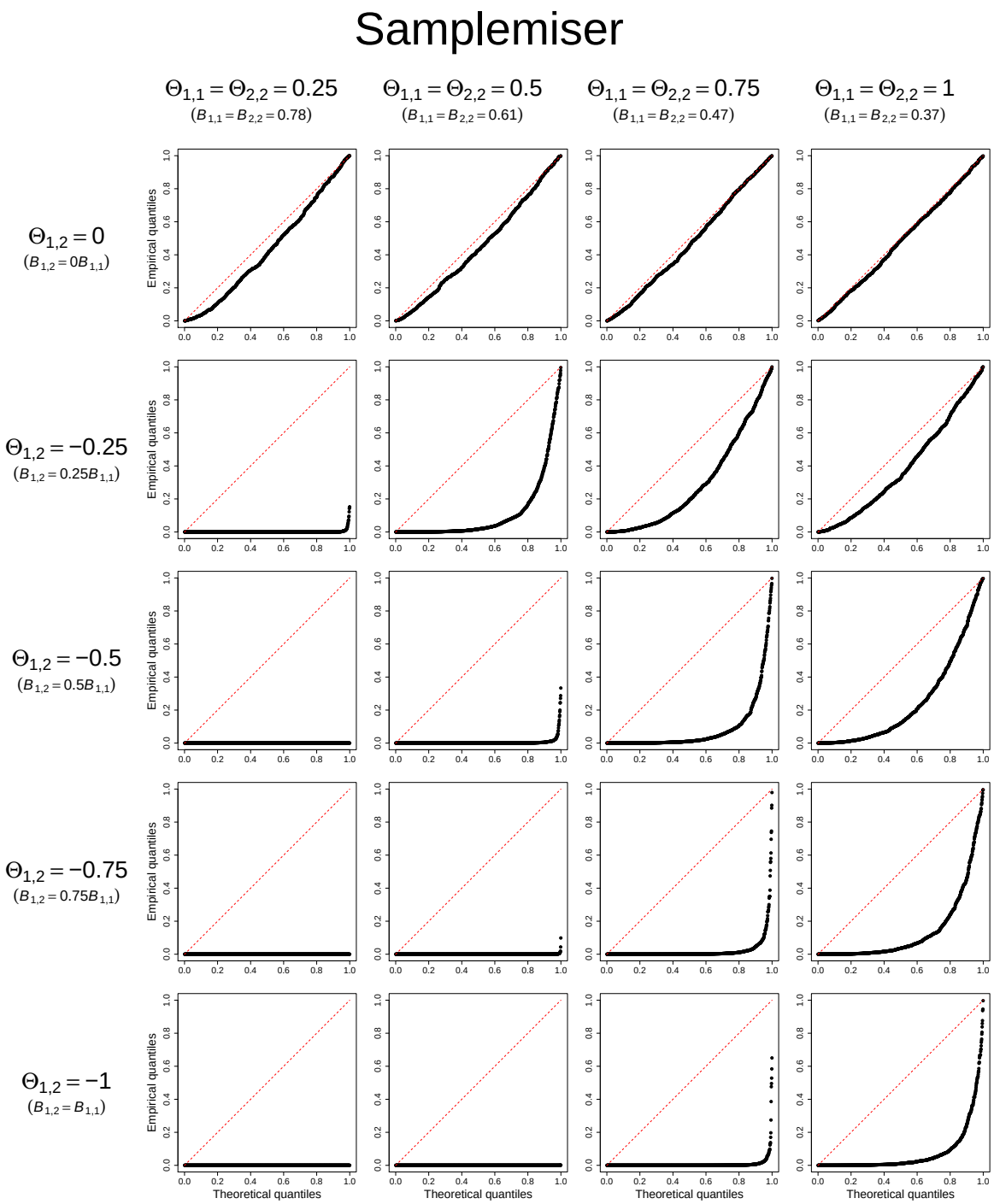
Note: 1,000 simulations per study. In each simulation, the first series consists of 320 randomly spaced instantaneous polls (averaging eight per year) and the second series consists of monthly instantaneous observations. The estimated parameter shown is $\mathbf{B}_{1,2}$, which equals zero under the data generating process. A dashed red line shows this true value of $\mathbf{B}_{1,2} = 0$ in each plot.

Figure A40: Histogram of distribution of estimates from Samplemiser and VAR models in imputation Monte Carlo study with average of twelve polls per year



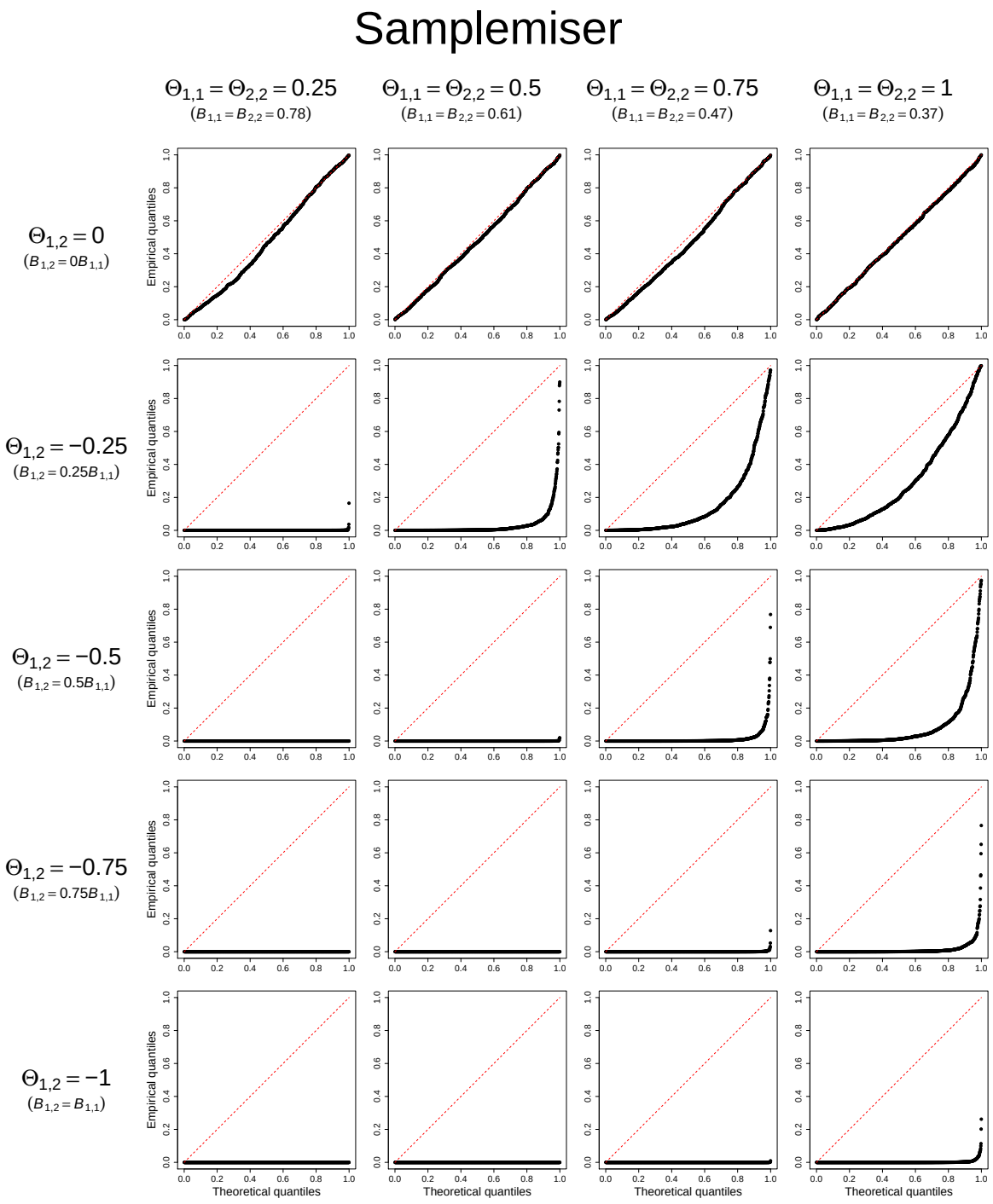
Note: 1,000 simulations per study. In each simulation, the first series consists of 480 randomly spaced instantaneous polls (averaging twelve per year) and the second series consists of monthly instantaneous observations. The estimated parameter shown is $\mathbf{B}_{1,2}$, which equals zero under the data generating process. A dashed red line shows this true value of $\mathbf{B}_{1,2} = 0$ in each plot.

Figure A41: Quantile-quantile plot of p-values from Samplemiser and VAR models in imputation Monte Carlo study with average of four polls per year



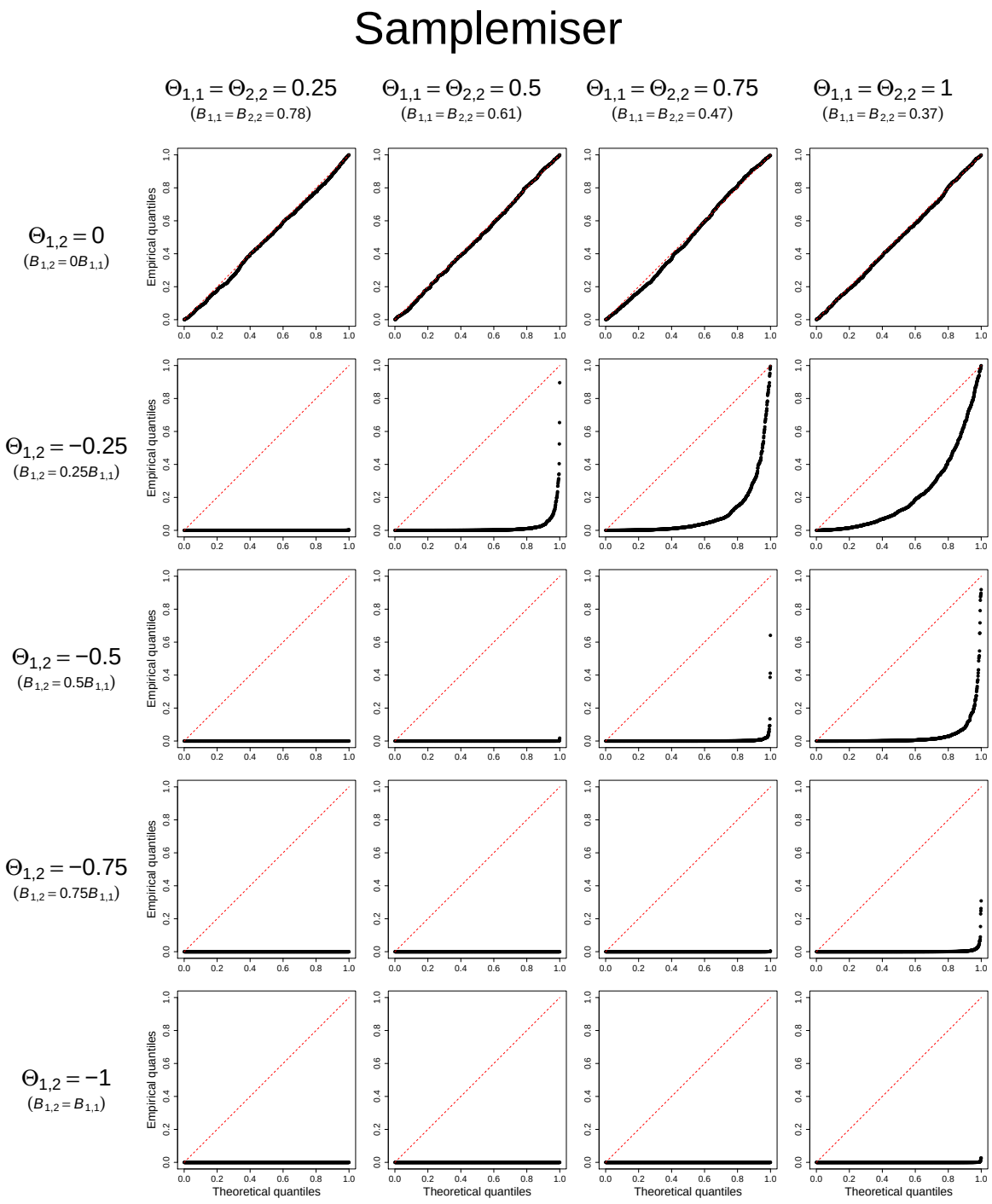
Note: 1,000 simulations per study. In each simulation, the first series consists of 160 randomly spaced instantaneous polls (averaging four per year) and the second series consists of monthly instantaneous observations. The estimated parameter shown is $\mathbf{B}_{1,2}$, which equals zero under the data generating process. The p-values correspond to a test of the null hypothesis that $\mathbf{B}_{1,2} = 0$. Since the null hypothesis is true under the data generating process, valid p-values should follow a uniform distribution.

Figure A42: Quantile-quantile plot of p-values from Samplemiser and VAR models in imputation Monte Carlo study with average of eight polls per year



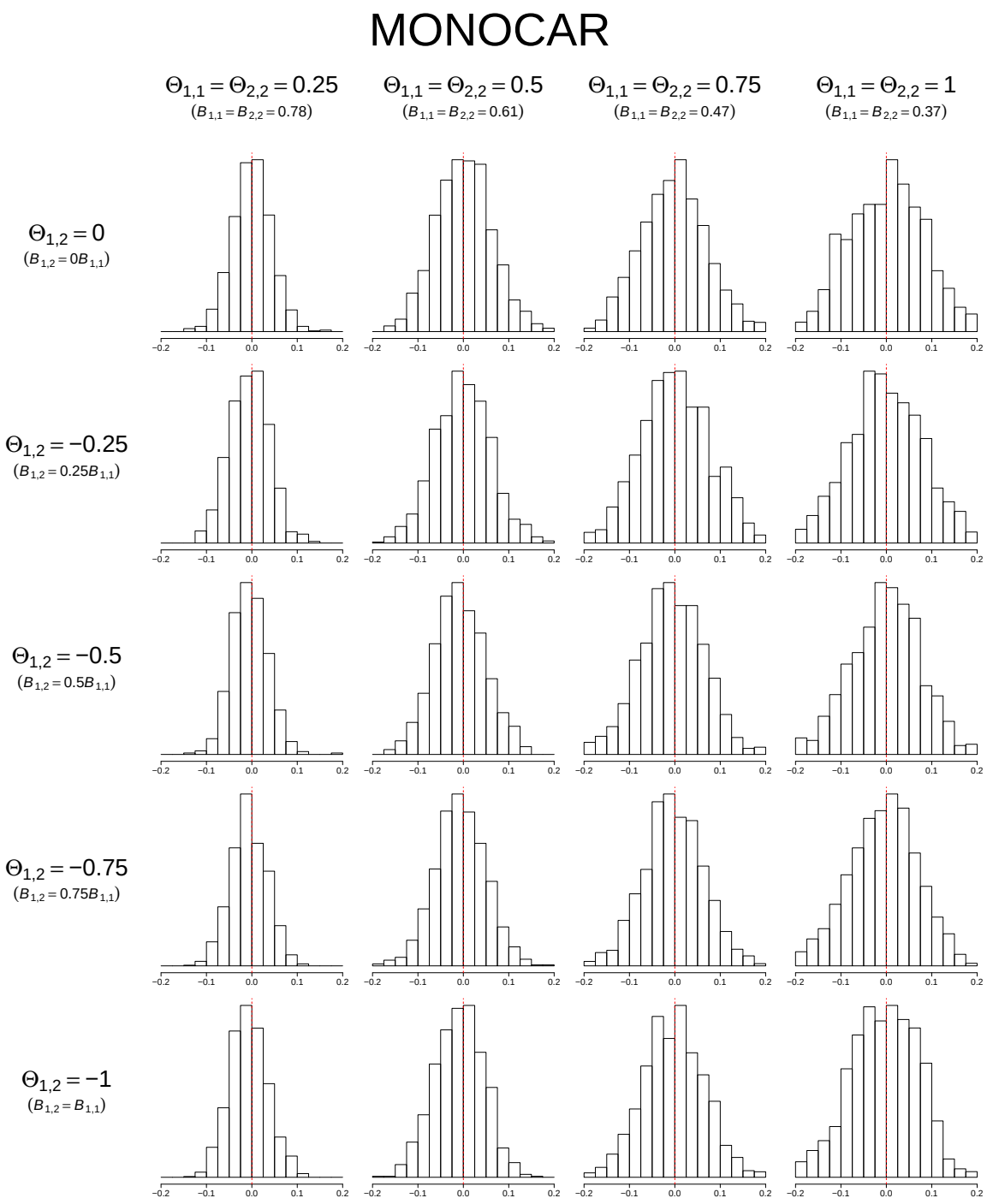
Note: 1,000 simulations per study. In each simulation, the first series consists of 320 randomly spaced instantaneous polls (averaging eight per year) and the second series consists of monthly instantaneous observations. The estimated parameter shown is $\mathbf{B}_{1,2}$, which equals zero under the data generating process. The p-values correspond to a test of the null hypothesis that $\mathbf{B}_{1,2} = 0$. Since the null hypothesis is true under the data generating process, valid p-values should follow a uniform distribution.

Figure A43: Quantile-quantile plot of p-values from Samplemiser and VAR models in imputation Monte Carlo study with average of twelve polls per year



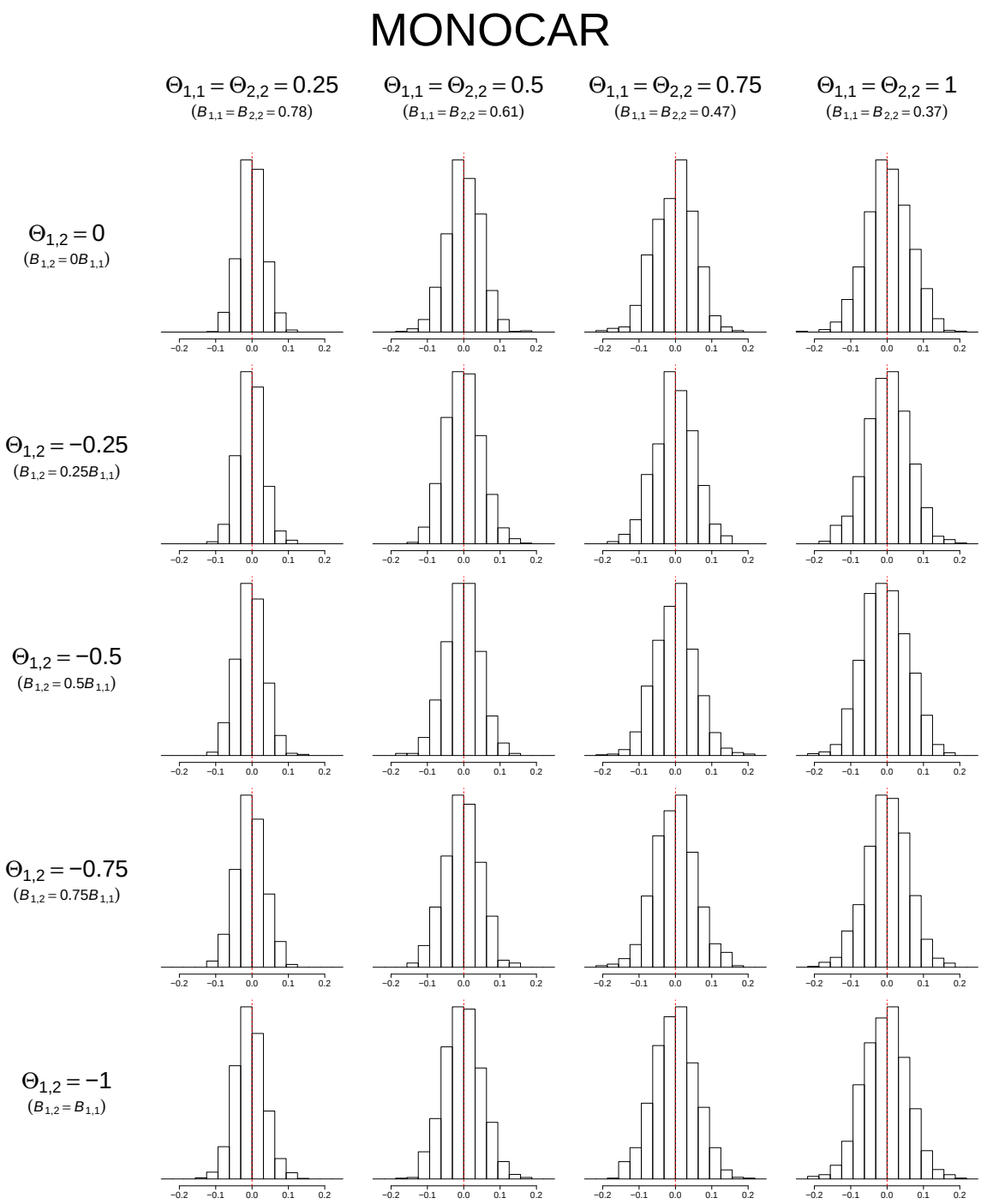
Note: 1,000 simulations per study. In each simulation, the first series consists of 480 randomly spaced instantaneous polls (averaging twelve per year) and the second series consists of monthly instantaneous observations. The estimated parameter shown is $\mathbf{B}_{1,2}$, which equals zero under the data generating process. The p-values correspond to a test of the null hypothesis that $\mathbf{B}_{1,2} = 0$. Since the null hypothesis is true under the data generating process, valid p-values should follow a uniform distribution.

Figure A44: Histogram of distribution of estimates from MONOCAR models in imputation Monte Carlo study with average of four polls per year



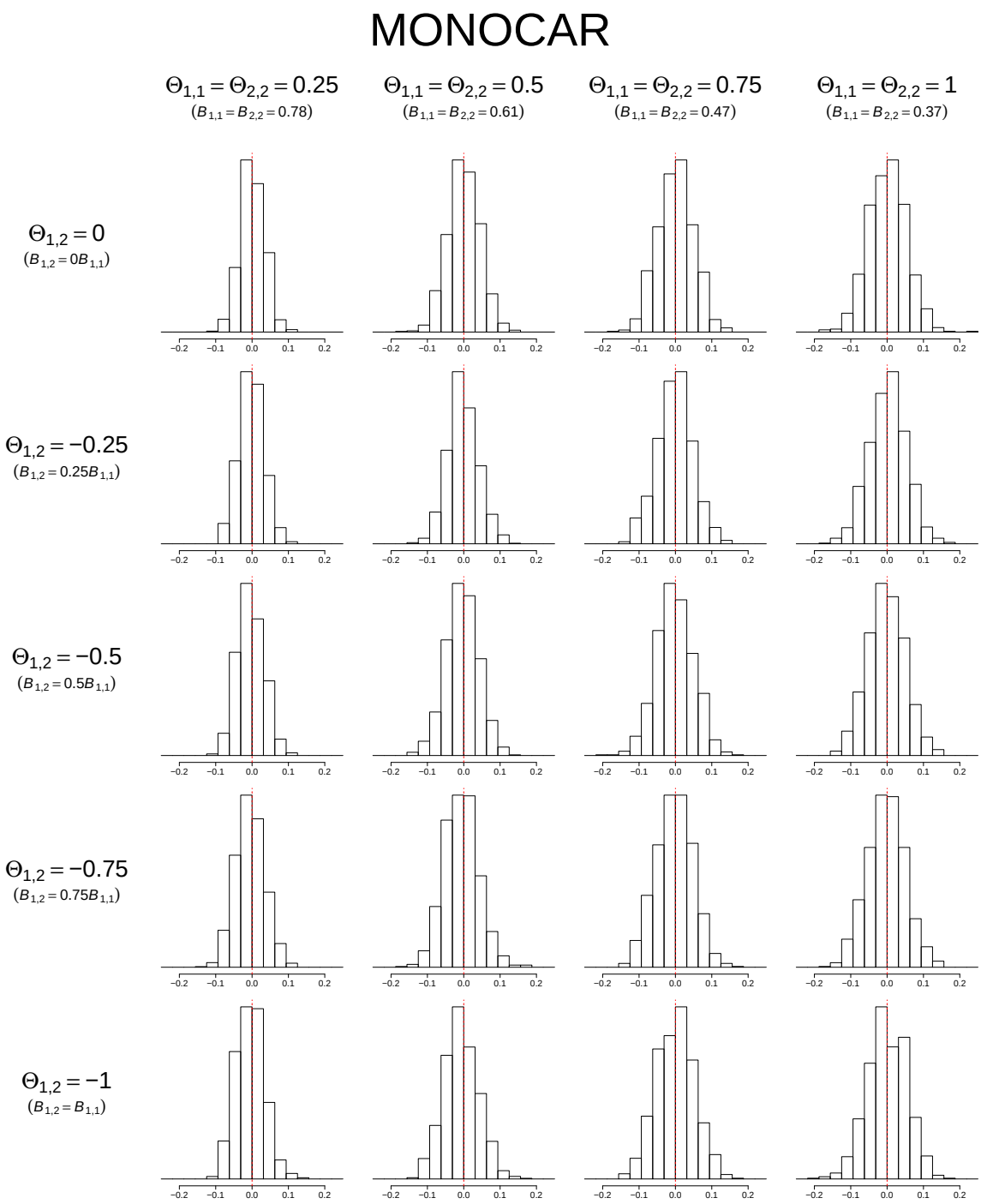
Note: 1,000 simulations per study. In each simulation, the first series consists of 160 randomly spaced instantaneous polls (averaging four per year) and the second series consists of monthly instantaneous observations. The estimated parameter shown is $\mathbf{B}_{1,2}$, which equals zero under the data generating process. A dashed red line shows this true value of $\mathbf{B}_{1,2} = 0$ in each plot.

Figure A45: Histogram of distribution of estimates from MONOCAR models in imputation Monte Carlo study with average of eight polls per year



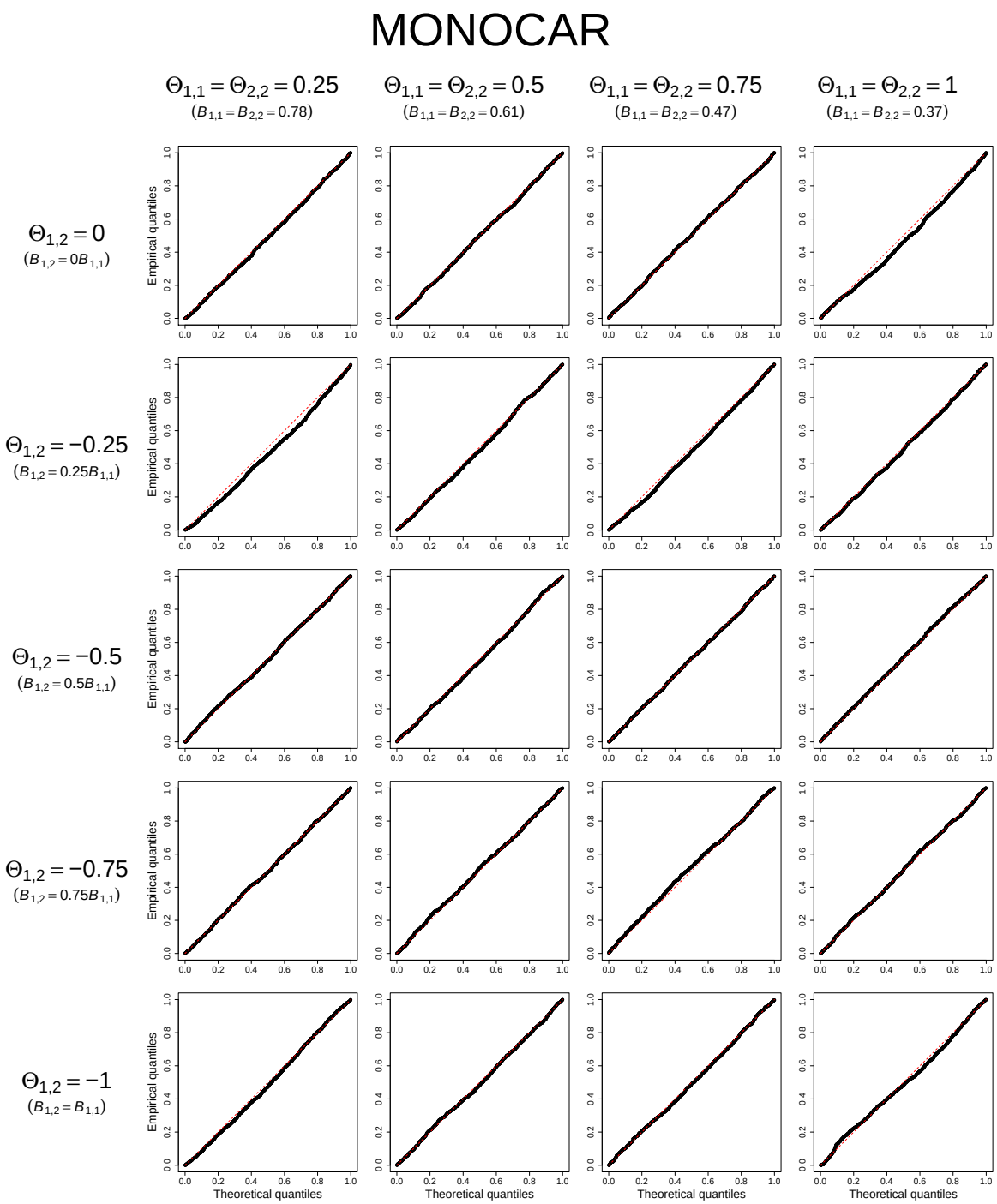
Note: 1,000 simulations per study. In each simulation, the first series consists of 320 randomly spaced instantaneous polls (averaging eight per year) and the second series consists of monthly instantaneous observations. The estimated parameter shown is $\mathbf{B}_{1,2}$, which equals zero under the data generating process. A dashed red line shows this true value of $\mathbf{B}_{1,2} = 0$ in each plot.

Figure A46: Histogram of distribution of estimates from MONOCAR models in imputation Monte Carlo study with average of twelve polls per year



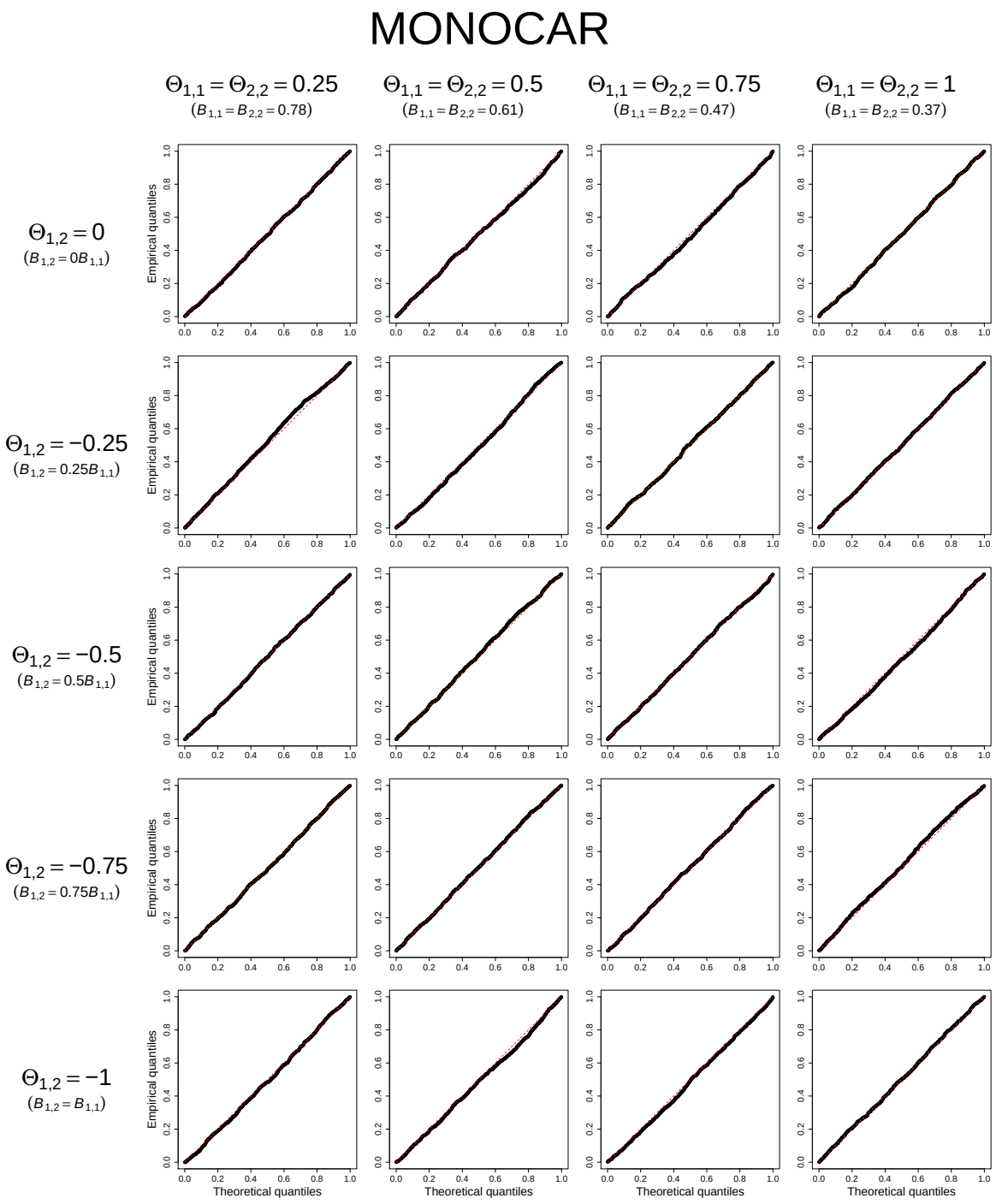
Note: 1,000 simulations per study. In each simulation, the first series consists of 480 randomly spaced instantaneous polls (averaging twelve per year) and the second series consists of monthly instantaneous observations. The estimated parameter shown is $\mathbf{B}_{1,2}$, which equals zero under the data generating process. A dashed red line shows this true value of $\mathbf{B}_{1,2} = 0$ in each plot.

Figure A47: Quantile-quantile plot of p-values from MONOCAR models in imputation Monte Carlo study with average of four polls per year



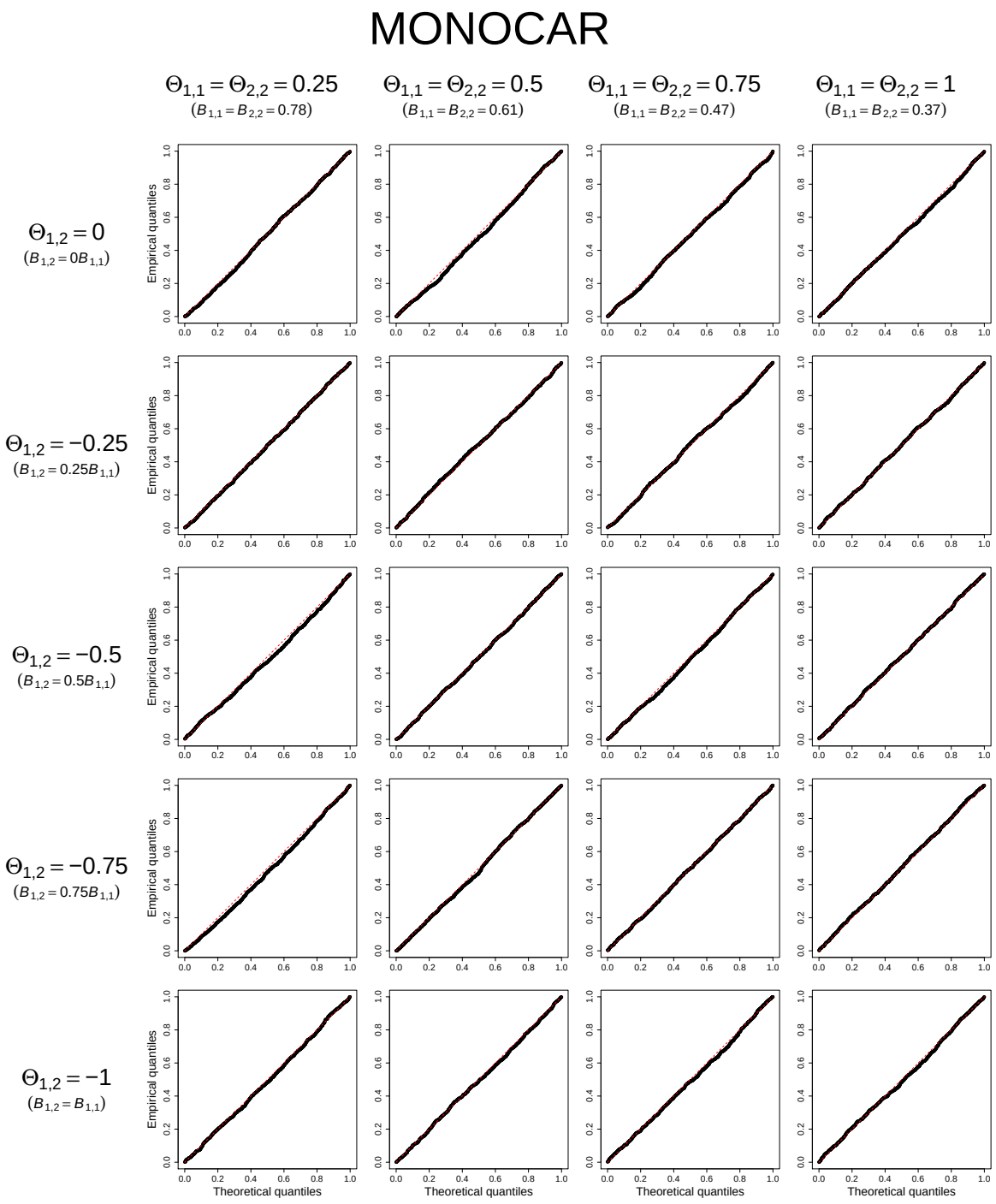
Note: 1,000 simulations per study. In each simulation, the first series consists of 160 randomly spaced instantaneous polls (averaging four per year) and the second series consists of monthly instantaneous observations. The estimated parameter shown is $\mathbf{B}_{1,2}$, which equals zero under the data generating process. The p-values correspond to a test of the null hypothesis that $\mathbf{B}_{1,2} = 0$. Since the null hypothesis is true under the data generating process, valid p-values should follow a uniform distribution.

Figure A48: Quantile-quantile plot of p-values from MONOCAR models in imputation Monte Carlo study with average of eight polls per year



Note: 1,000 simulations per study. In each simulation, the first series consists of 320 randomly spaced instantaneous polls (averaging eight per year) and the second series consists of monthly instantaneous observations. The estimated parameter shown is $\mathbf{B}_{1,2}$, which equals zero under the data generating process. The p-values correspond to a test of the null hypothesis that $\mathbf{B}_{1,2} = 0$. Since the null hypothesis is true under the data generating process, valid p-values should follow a uniform distribution.

Figure A49: Quantile-quantile plot of p-values from MONOCAR models in imputation Monte Carlo study with average of twelve polls per year



Note: 1,000 simulations per study. In each simulation, the first series consists of 480 randomly spaced instantaneous polls (averaging twelve per year) and the second series consists of monthly instantaneous observations. The estimated parameter shown is $\mathbf{B}_{1,2}$, which equals zero under the data generating process. The p-values correspond to a test of the null hypothesis that $\mathbf{B}_{1,2} = 0$. Since the null hypothesis is true under the data generating process, valid p-values should follow a uniform distribution.

I. COMPLETE MODEL ESTIMATES FOR PRESIDENTIAL APPROVAL EXAMPLE

Table A2: Full VAR/WCALC models of presidential approval and unemployment

	Model 7	Model 8	Model 11	Model 12
presidential approval				
approval_{t-1}	0.892*	0.892*	0.899*	0.899*
	(0.036)	(0.036)	(0.035)	(0.036)
$\text{unemployment}_{t-1}$	−0.507	−0.494	−0.046	−0.045
	(0.292)	(0.310)	(0.025)	(0.026)
Bush administration		0.107		0.007
		(0.841)		(0.070)
constant	9.692*	9.584*	0.322	0.314
	(3.623)	(3.734)	(0.179)	(0.198)
unemployment rate				
$\text{unemployment}_{t-1}$	0.997*	1.004*	1.000*	1.006*
	(0.012)	(0.013)	(0.012)	(0.013)
approval_{t-1}	0.001	0.001	0.020	0.017
	(0.001)	(0.001)	(0.017)	(0.017)
Bush administration		0.059		0.057
		(0.034)		(0.034)
constant	−0.049	−0.109	0.001	−0.067
	(0.148)	(0.151)	(0.088)	(0.096)
N (approval)	144	144	144	144
N (unemployment)	144	144	144	144
log-likelihood	−365.1	−363.5	−7.1	−5.6

Note: $^*p < 0.05$. Variance multiplier scaled to represent the multiplier for an observation with a sample size of 1000.

Table A3: Full MONOCAR models of presidential approval and unemployment

	– All polls –		– Gallup only –	
	Model 9	Model 10	Model 1	Model 2
presidential approval				
mean reversion rate of approval (Θ_{11})	0.302*	0.235*	0.245*	0.199*
	(0.082)	(0.073)	(0.081)	(0.075)
effect of unemployment on approval ($-\Theta_{21}$)	−0.024*	−0.026*	−0.022*	−0.024*
	(0.011)	(0.011)	(0.010)	(0.010)
shift in approval for Bush administration (δ_1)		−0.176*		−0.134
		(0.065)		(0.103)
long-run mean of approval (μ_1)	0.126	0.186	0.117	0.162
	(0.076)	(0.097)	(0.079)	(0.097)
variance multiplier for approval observations (ν_1)	3.433*	3.375*	4.048*	4.105*
	(0.290)	(0.283)	(0.712)	(0.699)
variance exponentiator for approval observations (ξ_1)	1.762*	1.734*	1.302*	1.333*
	(0.271)	(0.270)	(0.606)	(0.602)
unemployment rate				
mean reversion rate of unemployment (Θ_{22})	0.011	0.007	0.003	−0.001
	(0.013)	(0.014)	(0.013)	(0.014)
effect of approval on unemployment ($-\Theta_{12}$)	0.041	0.072	0.136	0.151
	(0.102)	(0.085)	(0.107)	(0.091)
shift in unemployment for Bush administration (δ_2)		−0.219		−0.225
		(0.168)		(0.166)
long-run mean of unemployment (μ_2)	7.218*	7.301*	7.225*	7.326*
	(0.825)	(0.800)	(0.817)	(0.783)
variance of unemployment observations (ν_2)	0.002	0.002	0.002	0.003
	(0.001)	(0.001)	(0.001)	(0.001)
N (approval)	533	533	212	212
N (unemployment)	144	144	144	144
log-likelihood	−1396.1	−1391.2	−678.0	−676.1

Note: $^*p < 0.05$. Variance multiplier scaled to represent the multiplier for an observation with a sample size of 1000.

Table A4: Full MARSS models of presidential approval and unemployment

	Model 3	Model 4	Model 5	Model 6
presidential approval				
approval$_{t-1}$ (B_{11})	0.890*	0.902*	0.994*	0.995*
	(0.038)	(0.034)	(0.002)	(0.002)
unemployment$_{t-1}$ (B_{21})	−0.096	−0.125*	−0.005	−0.006*
	(0.057)	(0.059)	(0.003)	(0.003)
Bush administration (δ_1)		−5.285*		−5.487*
		(2.013)		(2.470)
constant (α_1)	4.364	5.169	4.115	5.242
	(2.789)	(3.436)	(2.558)	(3.297)
variance multiplier for approval observations (ν_1)	6.481*	6.361*	2.484*	2.479*
	(0.443)	(0.436)	(0.224)	(0.221)
variance exponentiator for approval observations (ξ_1)	1.451*	1.435*	1.260*	1.258*
	(0.274)	(0.274)	(0.293)	(0.291)
unemployment rate				
unemployment$_{t-1}$ (B_{22})	0.993*	0.998*	1.000*	1.000*
	(0.011)	(0.012)	(0.000)	(0.000)
approval$_{t-1}$ (B_{12})	0.006	0.010	0.000	0.000
	(0.008)	(0.007)	(0.000)	(0.000)
Bush administration (δ_2)		−0.751		−0.723
		(0.546)		(0.551)
constant (α_2)	22.732*	23.570*	22.797*	23.541*
	(2.623)	(2.642)	(2.634)	(2.643)
variance of unemployment observations (ν_2)	0.000	0.000	0.000	0.000
	(0.023)	(0.022)	(0.024)	(0.023)
Time unit	Month		Day	
N (approval)	533	533	533	533
N (unemployment)	144	144	144	144
log-likelihood	−1472.7	−1468.1	−1353.2	−1349.5

Note: $^*p < 0.05$. Variance multiplier scaled to represent the multiplier for an observation with a sample size of 1000.

Table A5: Full MARSS models of presidential approval and unemployment, Gallup only

	Model 7	Model 8	Model 9	Model 10
presidential approval				
approval$_{t-1}$ (B_{11})	0.860*	0.881*	0.992*	0.993*
	(0.043)	(0.040)	(0.002)	(0.002)
unemployment$_{t-1}$ (B_{21})	−0.144*	−0.166*	−0.007*	−0.008*
	(0.063)	(0.065)	(0.003)	(0.003)
Bush administration (δ_1)		−4.808		−4.505
		(3.145)		(3.150)
constant (α_1)	3.697	5.246	3.656	5.124
	(2.560)	(3.067)	(2.471)	(3.025)
variance multiplier for approval observations (ν_1)	6.948*	6.930*	1.395*	1.429*
	(0.827)	(0.826)	(0.389)	(0.392)
variance exponentiator for approval observations (ξ_1)	1.440*	1.471*	2.714*	2.671*
	(0.610)	(0.609)	(1.108)	(1.078)
unemployment rate				
unemployment$_{t-1}$ (B_{22})	1.000*	1.004*	1.000*	1.000*
	(0.012)	(0.012)	(0.000)	(0.000)
approval$_{t-1}$ (B_{12})	0.015	0.017*	0.000	0.001*
	(0.009)	(0.008)	(0.000)	(0.000)
Bush administration (δ_2)		−0.748		−0.690
		(0.535)		(0.540)
constant (α_2)	22.807*	23.180*	22.880*	23.201*
	(2.563)	(2.396)	(2.614)	(2.514)
variance of unemployment observations (ν_2)	0.000	0.000	0.000	0.000
	(0.023)	(0.022)	(0.023)	(0.022)
Time unit	Month		Day	
N (approval)	212	212	212	212
N (unemployment)	144	144	144	144
log-likelihood	−698.0	−695.6	−651.8	−649.7

Note: $^*p < 0.05$. Variance multiplier scaled to represent the multiplier for an observation with a sample size of 1000.

Table A6: Lagged dependent variable/WCALC models of presidential approval

	– All polls –		– Gallup only –	
	Model 11	Model 12	Model 13	Model 14
approval_{t-1}	0.891*	0.890*	0.900*	0.899*
	(0.036)	(0.036)	(0.035)	(0.035)
unemployment_t	−0.541	−0.530	−0.047	−0.046
	(0.289)	(0.305)	(0.025)	(0.026)
Bush administration		0.104		0.009
		(0.833)		(0.069)
constant	10.017*	9.923*	0.328	0.318
	(3.578)	(3.669)	(0.177)	(0.194)
N (approval)	144	144	144	144
N (unemployment)	144	144	144	144
log-likelihood	−413.0	−413.0	−55.0	−55.0

Note: $^*p < 0.05$. Number of approval observations reflects number of imputed observations from WCALC, not number of polls used by WCALC.

References

Aguiar-Conraria, Luís, Pedro C. Magalhães, and Maria Joana Soares. 2012. "Cycles in Politics: Wavelet Analysis of Political Time Series." *American Journal of Political Science* 56 (2): 500–518.

Akaike, Hirotugu. 1978. "Covariance Matrix Computation of the State Variable of a Stationary Gaussian Process." *Annals of the Institute of Statistical Mathematics* 30 (1): 499–504.

Anderson, Brian D.O., and John B. Moore. 1979. *Optimal Filtering*. Englewood Cliffs, N.J.: Prentice-Hall.

Angelini, Elena, Jérôme Henry, and Massimiliano Marcellino. 2006. "Interpolation and Backdating with a Large Information Set." *Journal of Economic Dynamics and Control* 30 (December): 2693–2724.

Anscombe, Francis J. 1948. "The Transformation of Poisson, Binomial and Negative-Binomial Data." *Biometrika* 35 (3/4): 246–254.

Armstrong, II, David Alan. 2009. "Measuring the Democracy–Repression Nexus." *Electoral Studies* 28 (3): 403–412.

Baffigi, Alberto, Roberto Golinelli, and Giuseppe Parigi. 2004. "Bridge Models to Forecast the Euro Area GDP." *International Journal of Forecasting* 20 (3): 447–460.

Baltagi, Badi H., and Ping X. Wu. 1999. "Unequally Spaced Panel Data Regressions with AR(1) Disturbances." *Econometric Theory* 15 (06): 814–823.

Baltagi, Badi H., and Young-Jae Chang. 1994. "Incomplete panels." *Journal of Econometrics* 62 (June): 67–89.

Barning, Fredericus J.M. 1963. "The Numerical Analysis of the Light-Curve of 12 Lacertae." *Bulletin of the Astronomical Institutes of the Netherlands* 17 (1): 22–28.

Beck, Nathaniel. 1989. "Estimating Dynamic Models Using Kalman Filtering." *Political Analysis* 1 (1): 121–156.

Berardo, Ramiro, and John T. Scholz. 2010. "Self-Organizing Policy Networks: Risk, Partner Selection, and Cooperation in Estuaries." *American Journal of Political Science* 54 (3): 632–649.

Bergstrom, Albert Rex. 1988. "The History of Continuous-Time Econometric Models." *Econometric Theory* 4 (12): 365–383.

———. 1990. *Continuous Time Econometric Modelling*. Oxford: Oxford University Press.

———. 1996. "Survey of Continuous-Time Econometrics." In *Dynamic Disequilibrium Modeling: Theory and Applications: Proceedings of the Ninth International Symposium in Economic Theory and Econometrics*, ed. William A. Barnett, Giancarlo Gandolfo, and Claude Hillinger. International Symposia in Economic Theory and Econometrics Cambridge: Cambridge University Press.

Bond, Jon R., Richard Fleisher, and B. Dan Wood. 2003. "The Marginal and Time-Varying Effect of Public Approval on Presidential Success in Congress." *Journal of Politics* 65 (February): 92–110.

Brandt, Patrick T., and John R. Freeman. 2006. "Advances in Bayesian Time Series Modeling and the Study of Politics: Theory Testing, Forecasting, and Policy Analysis." *Political Analysis* 14 (1): 1–36.

———. 2009. "Modeling Macro-Political Dynamics." *Political Analysis* 17 (2): 113–142.

Brandt, Patrick T., and John T. Williams. 2001. "A Linear Poisson Autoregressive Model: The Poisson AR(p) Model." *Political Analysis* 9 (January): 164–184.

Brandt, Patrick T., John T. Williams, Benjamin O. Fordham, and Brain Pollins. 2000.

"Dynamic Modeling for Persistent Event-Count Time Series." *American Journal of Political Science* 44 (October): 823.

Brockwell, Peter J. 1994. "On Continuous-Time Threshold ARMA Processes." *Journal of Statistical Planning and Inference* 39 (2): 291–303.

———. 2001a. "Continuous-Time ARMA Processes." In *Stochastic Processes: Theory and Methods*, ed. Dayanand N. Shanbhag, and C. Radhakrishna Rao. Vol. 19 of *Handbook of Statistics* Amsterdam: North Holland, 149–276.

———. 2001b. "Lévy-Driven CARMA Processes." *Annals of the Institute of Statistical Mathematics* 53 (1): 113–124.

Brockwell, Peter J., and Rob J. Hyndman. 1992. "On Continuous-Time Threshold Autoregression." *International Journal of Forecasting* 8 (2): 157–173.

Broersen, Piet M.T. 2008. "Time Series Models for Spectral Analysis of Irregular Data Far Beyond the Mean Data Rate." *Measurement Science and Technology* 19 (January): 015103.

Carpenter, Daniel P. 2003. "Why Do Bureaucrats Delay? Lessons from a Stochastic Optimal Stopping Model of Agency Timing, with Applications to the FDA." In *Politics, Policy, and Organizations: Frontiers in the Study of Bureaucracy*, ed. George A. Krause, and Kenneth J. Meier. Ann Arbor: University of Michigan Press, 23–40.

Chanter, Dennis O. 1975. "Modifications of the Angular Transformation." *Journal of the Royal Statistical Society, Series C (Applied Statistics)* 24 (3): 354–359.

Chiu, Ching Wai Jeremy, Bjørn Eraker, Andrew T Foerster, Tae Bong Kim, and Hernán D Seoane. 2011. Estimating VARs Sampled at Mixed or Irregular Spaced Frequencies: A Bayesian Approach. Technical Report RWP 11-11 Federal Reserve Bank of Kansas City Kansas City: .

Clements, Michael P., and Ana Beatriz Galvão. 2008. "Macroeconomic Forecasting With Mixed-Frequency Data." *Journal of Business & Economic Statistics* 26 (4): 546–554.

Commandeur, Jacques J.F., and Siem Jan Koopman. 2007. *An Introduction to State Space Time Series Analysis*. Oxford: Oxford University Press.

De Boef, Suzanna, and Luke Keele. 2008. "Taking Time Seriously." *American Journal of Political Science* 52 (January): 184–200.

De Jong, Piet. 1988. "The Likelihood for a State Space Model." *Biometrika* 75 (1): 165–169.

Diron, Marie. 2008. "Short-term Forecasts of Euro Area Real GDP Growth: An Assessment of Real-time Performance Based on Vintage Data." *Journal of Forecasting* 27 (5): 371–390.

Dufour, Jean-Marie, and Marcel G. Dagenais. 1985. "Durbin-Watson tests for Serial Correlation in Regressions with Missing Observations." *Journal of Econometrics* 27 (March): 371–381.

Durbin, James, and Siem Jan Koopman. 2012. *Time Series Analysis by State Space Methods*. Oxford: Oxford University Press.

Erdogan, Emre, Sheng Ma, Alina Beygelzimer, and Irina Rish. 2005. "Statistical Models for Unequally Spaced Time Series." In *Proceedings of the 2005 SIAM International Conference on Data Mining*. SIAM, 626–630.

Fearon, James D. 1994. "Domestic Political Audiences and the Escalation of International Disputes." *American Political Science Review* 88 (03): 577–592.

———. 1998. "Bargaining, Enforcement, and International Cooperation." *International Organization* 52 (02): 269–305.

Fischer, Manuel, Karin Ingold, Pascal Sciarini, and Fr/'ed/'eric Varone. 2012. "Impacts of Market Liberalization on Regulatory Network: A Longitudinal Analysis of the Swiss Telecommunications Sector." *Policy Studies Journal* 40 (3): 435–457.

Foroni, Claudia, and Massimiliano Marcellino. 2013. A Survey of Econometric Methods for Mixed-frequency Data. Technical Report ECO 2013/02 Department of Economics, European University Institute Florence: .

Foroni, Claudia, Massimiliano Marcellino, and Christian Schumacher. 2012. U-MIDAS: MIDAS Regressions with Unrestricted Lag Polynomials. CEPR discussion paper CEPR.

Fuller, Wayne A. 1976. *Introduction to Statistical Time Series*. John Wiley & Sons.

Gardner, G., Andrew C. Harvey, and Garry D. A. Phillips. 1980. "Algorithm AS 154: An Algorithm for Exact Maximum Likelihood Estimation of Autoregressive-Moving Average Models by Means of Kalman Filtering." *Journal of the Royal Statistical Society, Series C (Applied Statistics)* 29 (3): 311–322.

Ghysels, Eric, Pedro Santa-Clara, and Rossen Valkanov. 2004. The MIDAS Touch: Mixed Data Sampling Regression Models. CIRANO Working Papers 2004s-20 CIRANO.

Green, Donald P., Alan S. Gerber, and Suzanna L. de Boef. 1999. "Tracking Opinion over Time: A Method for Reducing Sampling Error." *Public Opinion Quarterly* 63 (Summer): 178–192.

Gregory, Jr, Stanford W., and Timothy J. Gallagher. 2002. "Spectral Analysis of Candidates' Nonverbal Vocal Communication: Predicting U.S. Presidential Election Outcomes." *Social Psychology Quarterly* 65 (3): 298–308.

Hänggi, Peter, and Fabio Marchesoni. 2005. "Introduction: 100 Years of Brownian Motion." *Chaos* 15 (2): 026101.

Harvey, A. C., and G. D. A. Phillips. 1979. "Maximum Likelihood Estimation of Regression Models with Autoregressive-moving Average Disturbances." *Biometrika* 66 (1): 49–58.

Hoffman, Matthew D., and Andrew Gelman. 2014. "The No-U-Turn Sampler: Adaptively

Setting Path Lengths in Hamiltonian Monte Carlo." *Journal of Machine Learning Research* 15: 1351–1381.

Humpherys, Jeffrey, Preston Redd, and Jeremy West. 2012. "A Fresh Look at the Kalman Filter." *SIAM Review* 54 (4): 801–823.

Im, Eric Iksoon, Jon Cauley, and Todd Sandler. 1987. "Cycles and Substitutions in Terrorist Activities: A Spectral Approach." *Kyklos* 40 (2): 238–255.

Jackman, Simon. 2005. "Pooling the Polls over an Election Campaign." *Australian Journal of Political Science* 40 (December): 499–517.

Jones, Richard H. 1981. "Fitting a Continuous Time Autoregression to Discrete Data." In *Applied Time Series Analysis II*, ed. David F. Findley. New York: Academic Press, 651–682.

———. 1984. "Fitting Multivariate Models to Unequally Spaced Data." In *Time Series Analysis of Irregularly Observed Data*, ed. Emanuel Parzen. Vol. 25 of *Lecture Notes in Statistics* Springer New York, 158–188.

———. 1985. "Time Series Analysis with Unequally Spaced Data." In *Time Series in the Time Domain*, ed. Edward J. Hannan, Paruchuri R. Krishnaiah, and Malempati M. Rao. Vol. 5 of *Handbook of Statistics* Amsterdam: North Holland, 157–178.

Jones, Richard H., and Francis Boadi-Boateng. 1991. "Unequally Spaced Longitudinal Data with AR(1) Serial Correlation." *Biometrics* 47 (1): 161–175.

Jones, Richard H., and Lynn M. Ackerson. 1990. "Serial Correlation in Unequally Spaced Longitudinal Data." *Biometrika* 77 (4): 721–731.

Jones, Richard H., and Peter V. Tryon. 1987. "Continuous Time Series Models for Unequally Spaced Data Applied to Modeling Atomic Clocks." *SIAM Journal on Scientific and Statistical Computing* 8 (1): 71–81.

Kalman, Rudolph Emil. 1960. "A New Approach to Linear Filtering and Prediction Problems." *Transactions of the ASME–Journal of Basic Engineering* 82 (Series D): 35–45.

Kellstedt, Paul, E. McAvoy, and James A. Stimson. 1993. "Dynamic Analysis with Latent Constructs." *Political Analysis* 5 (1): 113–150.

Koopman, Siem Jan. 1997. "Exact Initial Kalman Filtering and Smoothing for Nonstationary Time Series Models." *Journal of the American Statistical Association* 92 (440): 1630–1638.

Lanning, Steven G. 1986. "Missing Observations: A Simultaneous Approach versus Interpolation by Related Series." *Journal of Economic and Social Measurement* 14 (2): 155–163.

Lévy-Leduc, C., E. Moulines, and F. Roueff. 2008. "Frequency Estimation Based on the Cumulated Lomb-Scargle Periodogram." *Journal of Time Series Analysis* 29 (6): 1104–1131.

Lomb, Nicholas R. 1976. "Least-squares Frequency Analysis of Unequally Spaced Data." *Astrophysics and space science* 39 (February): 447–462.

Manger, Mark S., Mark A. Pickup, and Tom A. B. Snijders. 2012. "A Hierarchy of Preferences: A Longitudinal Network Analysis Approach to PTA Formation." *Journal of Conflict Resolution* 56 (5): 853–878.

Mariano, Roberto S., and Yasutomo Murasawa. 2010. "A Coincident Index, Common Factors, and Monthly Real GDP." *Oxford Bulletin of Economics and Statistics* 72 (1): 27–46.

Marquardt, Tina, and Robert Stelzer. 2007. "Multivariate CARMA Processes." *Stochastic Processes and their Applications* 117: 96–120.

Martin, Andrew D., and Kevin M. Quinn. 2002. "Dynamic Ideal Point Estimation via Markov Chain Monte Carlo for the U.S. Supreme Court, 1953–1999." *Political Analysis* 10 (2): 134–153.

Matyasovszky, István. 2013. "Spectral Analysis of Unevenly Spaced Climatological Time Series." *Theoretical and Applied Climatology* 111 (3-4): 371–378.

McAvoy, Gregory E. 1998. "Partisan Probing and Democratic Decisionmaking Rethinking the Nimby Syndrome." *Policy Studies Journal* 26 (2): 274–292.

Mittnik, Stefan, and Peter Zadrozny. 2005. "Forecasting Quarterly German GDP at Monthly Intervals Using Monthly Ifo Business Conditions Data." In *Ifo Survey Data in Business Cycle and Monetary Policy Analysis*, ed. Jan-Egbert Sturm, and Timo Wollmershäuser. Contributions to Economics Physica-Verlag HD, 19–48.

Pickup, Mark A., and Christopher Wlezien. 2009. "On Filtering Longitudinal Public Opinion Data." *Electoral Studies* 28 (3): 354–367.

Pickup, Mark A., and Richard Johnston. 2007. "Campaign Trial Heats as Electoral Information." *Electoral Studies* 26 (2): 460–476.

Robinson, Peter M. 1985. "Testing for Serial Correlation in Regression with Missing Observations." *Journal of the Royal Statistical Society, Series B (Methodological)* 47 (3): 429–437.

Ruf, Thomas. 2010. "The Lomb-Scargle Periodogram in Biological Rhythm Research: Analysis of Incomplete and Unequally Spaced Time-Series." *Biological Rhythm Research* 30 (August): 178–201.

Ryan, Kevin F., and David E. A. Giles. 1998. "Testing for Unit Roots With Missing Observations." In *Messy Data: Missing Observations, Outliers, and Mixed-Frequency Data*, ed. Thomas B. Fomby, and R. Carter Hill. Vol. 13 of *Advances in Econometrics* JAI Press, 203–242.

Sandler, Todd, and Walter Enders. 2004. "An economic perspective on transnational terrorism." *European Journal of Political Economy* 20 (2): 301 – 316.

Savin, N. E., and Kenneth J. White. 1978. "Testing for Autocorrelation with Missing Observations." *Econometrica* 46 (1): 59–67.

Scargle, Jeffrey D. 1982. "Studies in Astronomical Time Series Analysis. II–Statistical Aspects of Spectral Analysis of Unevenly Spaced Data." *The Astrophysical Journal* 263 (December): 835–853.

Schlemm, Eckhard, and Robert Stelzer. 2012. "Multivariate CARMA Processes, Continuous-Time State Space Models and Complete Regularity of the Innovations of the Sampled Processes." *Bernoulli* 18 (1): 46–63.

Schulz, M, and K Stattegger. 1997. "SPECTRUM: Spectral Analysis of Unevenly Spaced Paleoclimatic Time Series." *Computers & Geosciences* 23 (9): 929–945.

Schweppe, Fred C. 1965. "Evaluation of Likelihood Functions for Gaussian Signals." *IEEE Transactions on Information Theory* 11 (1): 61–70.

Shively, Thomas S. 1993. "Testing for Autoregressive Disturbances in a Time Series Regression with Missing Observations." *Journal of Econometrics* 57 (May): 233–255.

Snijders, Tom A. B. 2001. "The Statistical Evaluation of Social Network Dynamics." *Sociological Methodology* 31 (1): 361–395.

Van det Eijk, Cees, and Robert Philip Weber. 1987. "Notes on the empirical analysis of cyclical processes." *European Journal of Political Research* 15 (2): 271–280.

Van Dongen, H. P. A., E. Olofsen, J. H. Van Hartevelt, and E. W. Kruyt. 2010. "A Procedure of Multiple Period Searching in Unequally Spaced Time-Series with the Lomb-Scargle Method." *Biological Rhythm Research* 30 (August): 149–177.

Vityazev, Veniamin V. 1996. "Time Series Analysis of Unequally Spaced Data: Intercomparison between the Schuster Periodogram and the LS-Spectra." *Astronomical & Astrophysical Transactions* 11 (2): 139–158.

Wansbeek, Tom, and Arie Kapteyn. 1985. "Estimation in a Linear Model with Serially Correlated Errors When Observations are Missing." *International Economic Review* 26 (2): 469–490.

———. 1989. "Estimation of the Error-Components Model with Incomplete Panels." *Journal of Econometrics* 41 (July): 341–361.

Yang, Yu. 2008. "Estimation for Lévy-driven CARMA Processes". Ph.D. diss. Colorado State University.

Zadrozny, Peter. 1988. "Gaussian Likelihood of Continuous-Time ARMAX Models When Data Are Stocks and Flows at Different Frequencies." *Econometric Theory* 4 (April): 108–124.