Supplementary Appendix to

Measuring Uncertainty about Long-Run Predictions

by Ulrich K. Müller and Mark W. Watson

1 Data Appendix

The data used in the paper are generally growth rates computed as differences in logarithms in percentage points at annual rate ($400 \times \ln(X_t/X_{t-1})$ if X_t is measured quarterly and $100 \times \ln(X_t/X_{t-1})$ if X_t is measured annually). The exceptions are stock returns which are $400 \times X_t$, where X_t is the gross quarterly real return (that is, $X_t = 1 + R_t$, where R_t is the net return), and TFP which is reported in percentage points at annual rate in the source data. The data are described in the following table.

Table B.1: Data Description

Series	Sample period	Description	Sources and Notes
		Post- WWII Quarterly Series	
GDP	1947:Q1-2014:Q4	Real GDP (Billions of Chained 2009 Dollars)	FRED Series: GDPC96
Consumption	1947:Q1-2014:Q4	Real Personal Consumption Expenditures (Billions of Chained 2009 Dollars)	FRED Series PCECC96
Total Factor Productivity	1947:Q2-2014:Q4	Growth Rate of TFP	From John Fernald's Web Page: Filename Quarterly) TFP.XLSX, series name DTFP. The series is described in Fernald (2014)
Labor Productivity	1947:Q1-2014:Q4	Output per hour in the non-farm business sector	FRED Series: OPHNFB
Population	1947:Q1-2014:Q4	Population (mid Quarter) (Thousands)	Bureau of Economic Analysis NIPA Table 7.1.
Prices: PCE Deflator	1947:Q1-2014:Q4	PCE deflator	FRED Series: PCECTPI
Inflation (CPI)	1947:Q1-2014:Q4	CPI	FRED Series: CPIAUCSL. The quarterly price index was computed as the average of the monthly values
Inflation (CPI, Japan)	1960:Q1-2014:Q4	CPI for Japan	FRED Series: CPI_JPN. The quarterly price index was computed as the average of the monthly values
Stock Returns	1947:Q1-2014:Q4	CRSP Real Value-Weighted Returns	CRSP Nominal Monthly Returns are from WRDS. Monthly real returns were computed by subtracting the change in the logarithm in the CPI from the nominal returns, which were then compounded to yield quarterly returns. Values shown are 400×the logarithm of gross quarterly real returns.
		Longer Span Data Series	
Real GDP	1900-2014	Real GDP (Billions of Chained 2005 Dollars)	1900-1929: Carter, Gartnter, Haines, Olmstead, Sutch, and Wright (2006), Table Ca9: Real GDP in $1996 1929-2011: BEA Real GDP in $2005. Data were linked in 1929
Real Consumption	1900-2014	Real Personal Consumption Expenditures (Billions of Chained 2005 Dollars)	1900-1929: Carter, Gartnter, Haines, Olmstead, Sutch, and Wright (2006), Table Cd78 Consumption expenditures, by type, Total, $1987 1929-2011: BEA Real Consumption in $2005. Data were linked in 1929
Population	1900-2014	Population	1900-1949: Carter, Gartnter, Haines, Olmstead, Sutch, and Wright (2006), Table Aa6 (Population Total including Armed Forces Oversees), 1917-1919 and 1930-1959; Table Aa7 (Total, Resident) 1900-1916 and 1920-1929) 1950-2011: Bureau of the Census (Total, Total including Armed Forces Oversees),
Inflation (CPI)	1913-2014	Consumer Price Index	FRED Series: CPIAUCNS. The annual price index was computed as the average of the monthly values
Stock Returns	1926:Q1-2014:Q4	CRSP Real Value-Weighted Returns	Described above
		Daily Data used for POOS forecasts	
Stock Returns	1926:M1:D2 – 2014:M12:D31	CRSP Nominal Value-Weighted Returns	CRSP nominal daily returns from WRDS.
3 Month Treasury Bill Rate	1954:M1:D4 – 2014:M12:D31	3-Month Treasury Bill Rates	FRED Series: DTB3

2 Proof of Theorem 1

We now introduce some notation that is used in the proof of Theorem 1, and in auxiliary Lemmas. We first state these Lemmas and then the proof of Theorem 1.

Notation: Define $\sigma_T^2 = \mathrm{Var}[T^{-1/2}\int_0^H g(s)x_{T,\lfloor sT\rfloor+1}ds]$ and $G(r) = \int_0^r g(s)ds$. Note that with $\tilde{g}_{T,t} = T\int_{(t-1)/T}^{t/T} g(s)ds$ and $\tilde{G}_{T,t} = T^{-1}\sum_{s=1}^{t-1}\tilde{g}_{T,t} = \sum_{s=1}^{t-1}\int_{(s-1)/T}^{s/T} g(s)ds = \int_0^{(t-1)/T} g(s)ds = G(\frac{t-1}{T})$, we find using summation by parts

$$\begin{aligned} T^{-1/2}\int_0^H g(s)x_{T,\lfloor sT\rfloor+1}ds &= T^{-3/2}\sum_{t=1}^{HT}\tilde{g}_{T,t}x_{T,t} \\ &= T^{-1/2}\tilde{G}_{T,HT+1}x_{T,HT} - T^{-1/2}\sum_{t=1}^{HT}\tilde{G}_T(t/T)\Delta x_{T,t} \\ &= -T^{-1/2}\sum_{t=1}^{HT} G(\frac{t-1}{T})\Delta x_{T,t} \end{aligned}$$

since $\tilde{G}_{T,HT+1} = G(H) = 0$.

Lemma B.1 *Under assumption (iv) of Theorem 1, there exists $C < \infty$ such that* $\sup_{0\le\lambda\le\pi,T}|\sum_{t=1}^{HT} e^{-\mathrm{i}\lambda t}G(\frac{t-1}{T})| \le C\lambda^{-2}T^{-1}$.

Proof. Let $r_t = \sum_{s=1}^{t} e^{\mathrm{i}\lambda s} = e^{\mathrm{i}\lambda}(1-e^{\mathrm{i}\lambda})^{-1}(e^{\mathrm{i}\lambda t}-1)$. By the above summation by parts argument,

$$\begin{aligned} \sum_{t=1}^{HT} e^{-\mathrm{i}\lambda t}G(\frac{t-1}{T}) &= -\int_0^H g(s)r_{\lfloor sT\rfloor+1}ds \\ &= -\frac{e^{\mathrm{i}\lambda}}{1-e^{\mathrm{i}\lambda}}\int_0^H g(s)e^{\mathrm{i}\lambda(\lfloor sT\rfloor+1)}ds \\ &= -\frac{e^{\mathrm{i}\lambda}}{1-e^{\mathrm{i}\lambda}}T^{-1}\sum_{t=1}^{HT}\tilde{g}_{T,t}e^{\mathrm{i}\lambda t} \end{aligned}$$

where the second equality uses $\int_0^H g(s)ds = 0$. Furthermore, by summation by parts,

$$\begin{aligned} \sum_{t=1}^{HT}\tilde{g}_{T,t}e^{\mathrm{i}\lambda t} &= \tilde{g}_{T,HT}r_{HT} - \sum_{t=1}^{HT}(\tilde{g}_{T,t}-\tilde{g}_{T,t-1})r_{t-1} \\ &= \frac{e^{\mathrm{i}\lambda}}{1-e^{\mathrm{i}\lambda}}\left(\tilde{g}_{T,HT}(e^{\mathrm{i}\lambda HT}-1) - \sum_{t=1}^{HT}(\tilde{g}_{T,t}-\tilde{g}_{T,t-1})(e^{\mathrm{i}\lambda(t-1)}-1)\right). \end{aligned}$$

Thus

$$\left|\sum_{t=1}^{HT} e^{-\mathrm{i}\lambda t}G(\frac{t-1}{T})\right| \le 2T^{-1}|e^{\mathrm{i}\lambda}(1-e^{\mathrm{i}\lambda})^{-1}|^2(\tilde{g}_{T,HT} + \sum_{t=1}^{HT}|\tilde{g}_{T,t}-\tilde{g}_{T,t-1}|).$$

For $0 \leq \lambda \leq \pi$, $|e^{\mathrm{i}\lambda}(1-e^{\mathrm{i}\lambda})^{-1}| = (2\sin(\lambda/2))^{-1} \leq 2\lambda$. Furthermore, since g is of bounded variation, it can be written as $g(s) = g^+(s) - g^-(s)$, with both g^+ and g^- positive and monotonically increasing. With $\tilde{g}^+_{T,t} = T\int_{(t-1)/T}^{t/T} g^+(s)ds$ and $\tilde{g}^-_{T,t} = T\int_{(t-1)/T}^{t/T} g^-(s)ds$, this implies the bound $\sum_{t=1}^{HT} |\tilde{g}_{T,t} - \tilde{g}_{T,t-1}| = \sum_{t=1}^{HT} |\tilde{g}^+_{T,t} - \tilde{g}^-_{T,t} - \tilde{g}^+_{T,t-1} + \tilde{g}^-_{T,t}| \leq \sum_{t=1}^{HT}(|\tilde{g}^+_{T,t} - \tilde{g}^+_{T,t-1}| + |\tilde{g}^-_{T,t} - \tilde{g}^-_{T,t-1}|) \leq g^+(HT) + g^-(HT) < \infty$, and the result follows. ■

Lemma B.2 *Under the assumptions Theorem 1,* $\sigma_T^2 \to \sigma^2 = \int_{-\infty}^{\infty} S(\omega) \left|\int_0^H e^{-\mathrm{i}\omega s} g(s)ds\right|^2 d\omega$.

Proof. Let γ_T be the autocovariances of $\Delta x_{T,t}$. We find

$$\begin{aligned}
\mathrm{Var}[T^{-1/2}\sum_{t=1}^{HT} G(\frac{t-1}{T})\Delta x_{T,t}] &= T^{-1}\sum_{j,k=1}^{HT} \gamma_T(k-j)G(\frac{k-1}{T})G(\frac{j-1}{T}) \\
&= T^{-1}\sum_{j,k=1}^{HT} \left(\int_{-\pi}^{\pi} e^{\mathrm{i}\lambda(k-j)} F_T(\lambda)d\lambda\right) G(\frac{k-1}{T})G(\frac{j-1}{T}) \\
&= T^{-1}\int_{-\pi}^{\pi} F_T(\lambda)\left|\sum_{t=1}^{HT} e^{\mathrm{i}\lambda t} G(\frac{t-1}{T})\right|^2 d\lambda.
\end{aligned}$$

Now for any fixed K, we will show

$$\begin{aligned}
T^{-1}\int_{-K/T}^{K/T} F_T(\lambda)\left|\sum_{t=1}^{HT} e^{\mathrm{i}\lambda t} G(\frac{t-1}{T})\right|^2 d\lambda &= \int_{-K}^{K} F_T(\frac{\omega}{T})\left|T^{-1}\sum_{t=1}^{HT} e^{\mathrm{i}\omega t/T} G(\frac{t-1}{T})\right|^2 d\omega \\
&\to \int_{-K}^{K} S(\omega)\omega^2\left|\int_0^H e^{\mathrm{i}\omega s} G(s)ds\right|^2 d\omega. \qquad \text{(B.1)}
\end{aligned}$$

First note that for any two complex numbers a, b, $|a-b| \geq ||a|-|b||$, so that $||a|^2 - |b|^2| = |(|a|+|b|)(|a|-|b|)| \leq (|a|+|b|)|a-b|$. Thus,

$$\left|\left|\int_0^H e^{\mathrm{i}\omega s} G(s)ds\right|^2 - \left|T^{-1}\sum_{t=1}^{HT} e^{\mathrm{i}\omega t/T} G(\frac{t-1}{T})\right|^2\right| \leq$$

$$2H\sup_s |G(s)| \cdot \left|T^{-1}\sum_{t=1}^{HT} e^{\mathrm{i}\omega t/T} G(\frac{t-1}{T}) - \int_0^H e^{\mathrm{i}\omega s} G(s)ds\right|$$

and

$$\begin{aligned}
&\left|T^{-1}\sum_{t=1}^{HT} e^{\mathrm{i}\omega t/T} G(\frac{t-1}{T}) - \int_0^H e^{\mathrm{i}\omega s} G(s)ds\right| \\
\leq &\int_0^H |e^{\mathrm{i}\omega s} G(s) - e^{\mathrm{i}\omega \lfloor sT\rfloor/T} G(\lfloor sT\rfloor/T)|ds \\
= &\int_0^H |e^{\mathrm{i}\omega s}(G(s) - G(\lfloor sT\rfloor/T)) + G(\lfloor sT\rfloor/T)(e^{\mathrm{i}\omega s} - e^{\mathrm{i}\omega\lfloor sT\rfloor/T})|ds
\end{aligned}$$

$$\leq \int_0^H |G(s) - G(\lfloor sT \rfloor/T)|ds + \sup_s |G(s)| \int_0^H |1 - e^{\mathrm{i}\omega(\lfloor sT \rfloor/T - s)}|ds.$$

Since for any real a, $|1-e^{\mathrm{i}a}| < |a|$, $\sup_{|\omega|\leq K} |1-e^{\mathrm{i}\omega(\lfloor sT \rfloor/T-s)}| \leq K/T$, and also $|G(s)-G(\lfloor sT \rfloor/T)| \leq T^{-1} \sup_s |g(s)|$. Thus,

$$\sup_{|\omega|\leq K} \left| \left| T^{-1} \sum_{t=1}^{HT} e^{\mathrm{i}\omega t/T} G(\frac{t-1}{T}) \right|^2 - \left| \int_0^H e^{\mathrm{i}\omega s} G(s) ds \right|^2 \right| \to 0$$

and (B.1) follows from assumption (iii.a) by straightforward arguments.

Further, since $\int_0^H e^{\mathrm{i}\omega s} g(s) ds = -\mathrm{i}\omega \int_0^H e^{\mathrm{i}\omega s} G(s) ds$,

$$\int_{-K}^{K} S(\omega)\omega^2 \left| \int_0^H e^{\mathrm{i}\omega s} G(s) ds \right|^2 d\omega = \int_{-K}^{K} S(\omega) \left| \int_0^H e^{\mathrm{i}\omega s} g(s) ds \right|^2 d\omega.$$

Thus, for any fixed K,

$$\delta_K(T) = T^{-1} \int_{-K/T}^{K/T} F_T(\lambda) \left| \sum_{t=1}^{HT} e^{\mathrm{i}\lambda t} G(\frac{t-1}{T}) \right|^2 d\lambda - \int_{-K}^{K} S(\omega) \left| \int_0^H e^{\mathrm{i}\omega s} g(s) ds \right|^2 d\omega \to 0.$$

Now for each T, define K_T as the largest integer $K \leq T$ for which $\sup_{T' \geq T} |\delta_K(T')| \leq 1/K$ (and zero if no such K exists). Note that $\delta_K(T) \to 0$ for all fixed K implies that $K_T \to \infty$, and by construction, also $\delta_{K_T}(T) \to 0$. The result now follows from

$$T^{-1} \int_{K_T/T}^{\pi} F_T(\lambda) \left| \sum_{t=1}^{HT} e^{\mathrm{i}\lambda t} G(\frac{t-1}{T}) \right|^2 d\lambda \leq C^2 T^{-3} \int_{K_T/T}^{\pi} F_T(\lambda) \frac{1}{\lambda^4} d\lambda \to 0$$

by assumption (iii.b) and Lemma B.1. ■

Lemma B.3 *Under the assumptions Theorem 1,* $T^{-1/2} \sup_{t,T} |\sum_{j=1}^{HT} G(\frac{j-1}{T}) c_{T,j-t}| \to 0$.

Proof. Recall that for any two real, square integrable sequences $\{a_j\}_{j=-\infty}^{\infty}$ and $\{b_j\}_{j=-\infty}^{\infty}$, $\sum_{j=-\infty}^{\infty} a_j b_j = \frac{1}{2\pi} \int_{-\pi}^{\pi} \hat{A}(\lambda)\hat{B}^*(\lambda) d\lambda$, where $\hat{A}(\lambda) = \sum_{j=-\infty}^{\infty} a_j e^{-\mathrm{i}\lambda j}$ and $\hat{B}^*(\lambda) = \sum_{j=-\infty}^{\infty} b_j e^{\mathrm{i}\lambda j}$. Thus

$$\sum_{j=1}^{HT} G(\frac{j-1}{T}) c_{T,j-t} = \sum_{j=1-t}^{HT-t} G(\frac{t+j-1}{T}) c_{T,j} = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{\mathrm{i}\lambda t} \hat{G}_T(\lambda) \hat{C}_T^*(\lambda) d\lambda$$

where $\hat{C}_T(\lambda) = \sum_{j=-\infty}^{\infty} c_{T,j} e^{-\mathrm{i}\lambda j}$ and $\hat{G}_T(\lambda) = \sum_{j=1}^{HT} e^{-\mathrm{i}\lambda j} G(\frac{j-1}{T})$, so that $e^{\mathrm{i}\lambda t}\hat{G}_T(\lambda) = \sum_{j=1}^{HT} e^{-\mathrm{i}\lambda(j-t)} G(\frac{t-1}{T}) = \sum_{j=1-t}^{HT-t} e^{-\mathrm{i}\lambda j} G(\frac{t+j-1}{T})$, and $\hat{C}_T^*$ is the complex conjugate of $\hat{C}_T$. Now since $F_T(\lambda) = \frac{1}{2\pi} |\hat{C}_T(\lambda)|^2$, we find

$$2\pi \left| \sum_{j=1}^{HT} G(\frac{j-1}{T}) c_{T,j-t} \right| = \left| \int_{-\pi}^{\pi} e^{\mathrm{i}\lambda t} \hat{G}_T(\lambda) \hat{C}_T^*(\lambda) d\lambda \right| \leq \sqrt{2\pi} \int_{-\pi}^{\pi} |\hat{G}_T(\lambda)| F_T(\lambda)^{1/2} d\lambda.$$

Also, using $|\hat{G}_T(\lambda)| \leq \sum_{j=1}^{HT} |G(\frac{j-1}{T})|$, we have

$$\begin{aligned}\int_0^{1/T} |\hat{G}_T(\lambda)| F_T(\lambda)^{1/2} d\lambda &\leq T^{-1} \sum_{j=1}^{HT} |G(\frac{j-1}{T})| \cdot \int_0^1 F_T(\omega/T)^{1/2} d\omega \\ &\to \int_0^H |G(s)| ds \cdot \int_0^1 \omega S(\omega)^{1/2} d\omega < \infty\end{aligned}$$

where the convergence follows from assumption (iii.a). Furthermore, by Lemma B.1,

$$\int_{1/T}^{\pi} |\hat{G}_T(\lambda)| F_T(\lambda)^{1/2} d\lambda \leq C T^{-1} \int_{1/T}^{\pi} F_T(\lambda)^{1/2} \lambda^{-2} d\lambda.$$

The result now follows from assumption (iii.c). ■

Lemma B.4 *Under the assumptions Theorem 1, for every $\epsilon > 0$ there exists a $M > 0$ such that*

$$\limsup_{T\to\infty} \mathrm{Var}[T^{-1/2} \int_0^H g(s) x_{T,\lfloor sT \rfloor + 1} ds + T^{-1/2} \sum_{t=-MT}^{MT} \left(\sum_{j=1}^{HT} G(\frac{j-1}{T}) c_{T,j-t} \right) \varepsilon_t] < \epsilon.$$

For this M, $\sigma^2_{T,M} = \mathrm{Var}[T^{-1/2} \sum_{t=-MT}^{MT} \left(\sum_{j=1}^{HT} G(\frac{j-1}{T}) c_{T,j-t} \right) \varepsilon_t]$ satisfies $\limsup_{T\to\infty} |\sigma^2_{M,T} - \sigma^2| < \epsilon$.

Proof. We have

$$\begin{aligned}-\int_0^H g(s) x_{T,\lfloor sT \rfloor + 1} ds &= \sum_{t=1}^{HT} G(\frac{t-1}{T}) \Delta x_{T,t} \\ &= \sum_{t=1}^{HT} G(\frac{t-1}{T}) \sum_{s=-\infty}^{\infty} c_{T,s} \varepsilon_{t-s} \\ &= \sum_{t=-\infty}^{\infty} \left(\sum_{j=1}^{HT} G(\frac{j-1}{T}) c_{T,j-t} \right) \varepsilon_t\end{aligned}$$

so that, with $\bar{G} = \sup_{0 \leq s < H} |G(s)|$

$$\begin{aligned}&\mathrm{Var}[T^{-1/2} \int_0^H g(s) x_{T,\lfloor sT \rfloor + 1} ds + T^{-1/2} \sum_{t=-MT}^{MT} \left(\sum_{j=1}^{HT} G(\frac{j-1}{T}) c_{T,j-t} \right) \varepsilon_t] \\ = \; & T^{-1} \sum_{t=-\infty}^{-MT-1} \left(\sum_{j=1}^{HT} G(\frac{j-1}{T}) c_{T,j-t} \right)^2 + T^{-1} \sum_{t=MT+1}^{\infty} \left(\sum_{j=1}^{HT} G(\frac{j-1}{T}) c_{T,j-t} \right)^2 \\ \leq \; & \bar{G}^2 T^{-1} \sum_{t=MT+1}^{\infty} \left(\sum_{j=1}^{HT} (|c_{T,j-t}| + |c_{T,j+t}|) \right)^2\end{aligned}$$

$$\leq \quad 4H^2\bar{G}^2T^{-1}\sum_{t=MT+1}^{\infty}\left(T\sup_{|s|\geq t-HT}|c_{T,s}|\right)^2$$
$$= \quad 4H^2\bar{G}^2T^{-1}\sum_{t=(M-H)T+1}^{\infty}\left(T\sup_{|s|\geq t}|c_{T,s}|\right)^2$$

which can be made arbitrarily small by choosing M large enough via assumption (ii).

The second claim follows directly from Lemma B.2. ■

Lemma B.5 *Under the assumptions Theorem 1 and if $\sigma^2 > 0$, then for all large enough integers $M > 0$, $\sigma_{M,T}^{-1}T^{-1/2}\sum_{t=-MT}^{MT}\left(\sum_{j=1}^{HT}G(\frac{j-1}{T})c_{T,j-t}\right)\varepsilon_t \Rightarrow \mathcal{N}(0,1)$.*

Proof. By the second claim in Lemma B.4 and Lemma B.2, $\sigma_{M,T} = O(1)$ and $\sigma_{M,T}^{-1} = O(1)$. By Theorem 24.3 in Davidson (1994), it thus suffices to show (a) $T^{-1/2}\sup_{1\leq t\leq HT}|\sum_{j=1}^{HT}G(\frac{j-1}{T})c_{T,j-t}\varepsilon_t| \xrightarrow{p} 0$ and (b) $T^{-1}\sum_{t=-MT}^{MT}\left(\sum_{j=1}^{HT}G(\frac{j-1}{T})c_{T,j-t}\right)^2(\varepsilon_t^2 - 1) \xrightarrow{p} 0$.

(a) is implied by the Lyapunov condition via Davidson's (1994) Theorems 23.16 and 23.11. Thus, it suffices to show that

$$\sum_{t=-MT}^{MT}E[|T^{-1/2}\left(\sum_{j=1}^{HT}G(\frac{j-1}{T})c_{T,j-t}\right)\varepsilon_t|^{2+\delta}] \to 0.$$

Now

$$\sum_{t=-MT}^{MT}E\left[\left|T^{-1/2}\left(\sum_{j=1}^{HT}G(\frac{j-1}{T})c_{T,j-t}\right)\varepsilon_t\right|^{2+\delta}\right]$$
$$\leq \quad T^{-1-\delta/2}\sum_{t=-MT}^{MT}\left|\sum_{j=1}^{HT}G(\frac{j-1}{T})c_{T,j-t}\right|^{2+\delta}E(|\varepsilon_t|^{2+\delta})$$
$$\leq \quad (\sup_t E(|\varepsilon_t|^{2+\delta}))\cdot T^{-\delta/2}\sup_t\left|\sum_{j=1}^{HT}G(\frac{j-1}{T})c_{T,j-t}\right|^{\delta}\cdot T^{-1}\sum_{t=-MT}^{MT}\left|\sum_{j=1}^{HT}G(\frac{j-1}{T})c_{T,j-t}\right|^2$$
$$= \quad (\sup_t E(|\varepsilon_t|^{2+\delta}))\cdot\left(T^{-1/2}\sup_t\left|\sum_{j=1}^{HT}G(\frac{j-1}{T})c_{T,j-t}\right|\right)^{\delta}\cdot\sigma_{M,T}^2 \to 0$$

where the convergence follows from Lemma B.3.

For (b), we apply Theorem 19.11 of Davidson (1994) with Davidson's X and c chosen as $X_{T,t}^D = c_{T,t}^D(\varepsilon_t^2 - 1)$ and $c_{T,t}^D = T^{-1}\left(\sum_{j=1}^{HT}G(\frac{j-1}{T})c_{T,j-t}\right)^2$. Then $X_{T,t}^D/c_{T,t}^D = \varepsilon_t^2 - 1$ is uniformly integrable, since $\sup_t E(|\varepsilon_t|^{2+\delta}) < \infty$. Further,

$$\sum_{t=-MT}^{MT}c_{T,t}^D = T^{-1}\sum_{t=-MT}^{MT}\left(\sum_{j=1}^{HT}G(\frac{j-1}{T})c_{T,j-t}\right)^2 = \sigma_{M,T}^2 = O(1)$$

and

$$\begin{aligned}\sum_{t=-MT}^{MT}(c_{T,t}^{D})^2 &= T^{-2}\sum_{t=-MT}^{MT}\left(\sum_{j=1}^{HT}G(\frac{j-1}{T})c_{T,j-t}\right)^4 \\ &\leq \sup_{-MT\leq t\leq MT}\left|T^{-1/2}\sum_{j=1}^{HT}G(\frac{j-1}{T})c_{T,j-t}\right|^2\cdot T^{-1}\sum_{t=-MT}^{MT}\left(\sum_{j=1}^{HT}G(\frac{j-1}{T})c_{T,j-t}\right)^2 \\ &= \sup_{-MT\leq t\leq MT}\left|T^{-1/2}\sum_{j=1}^{HT}G(\frac{j-1}{T})c_{T,j-t}\right|^2\cdot\sigma_{M,T}^2\to 0\end{aligned}$$

where the convergence follows from Lemma B.3. ■

Proof of Theorem 1:

If σ^2 in Lemma B.2 is zero, then $T^{-1/2}\int_0^H g(s)x_{T,\lfloor sT\rfloor+1}ds \xrightarrow{p} 0$ follows immediately from Lemma B.2 via mean square convergence. If $\sigma^2>0$, then by Lemmas B.2, B.4 and B.5, for large enough M,

$$\sigma^{-1}T^{-1/2}\int_0^H g(s)x_{T,\lfloor sT\rfloor+1}ds=\frac{\sigma_{M,T}}{\sigma}A_T+\frac{\sigma_{M,T}}{\sigma}B_T$$

where $A_T\Rightarrow\mathcal{N}(0,1)$, $\limsup_{T\to\infty}E(B_T^2)<\epsilon$ and $\limsup_{T\to\infty}|\sigma_{M,T}^2-\sigma^2|<\epsilon$. Thus, by Slutzky's Theorem, as $\epsilon\to 0$, the desired convergence in distribution follows. But ϵ was arbitrary, which proves the Theorem.

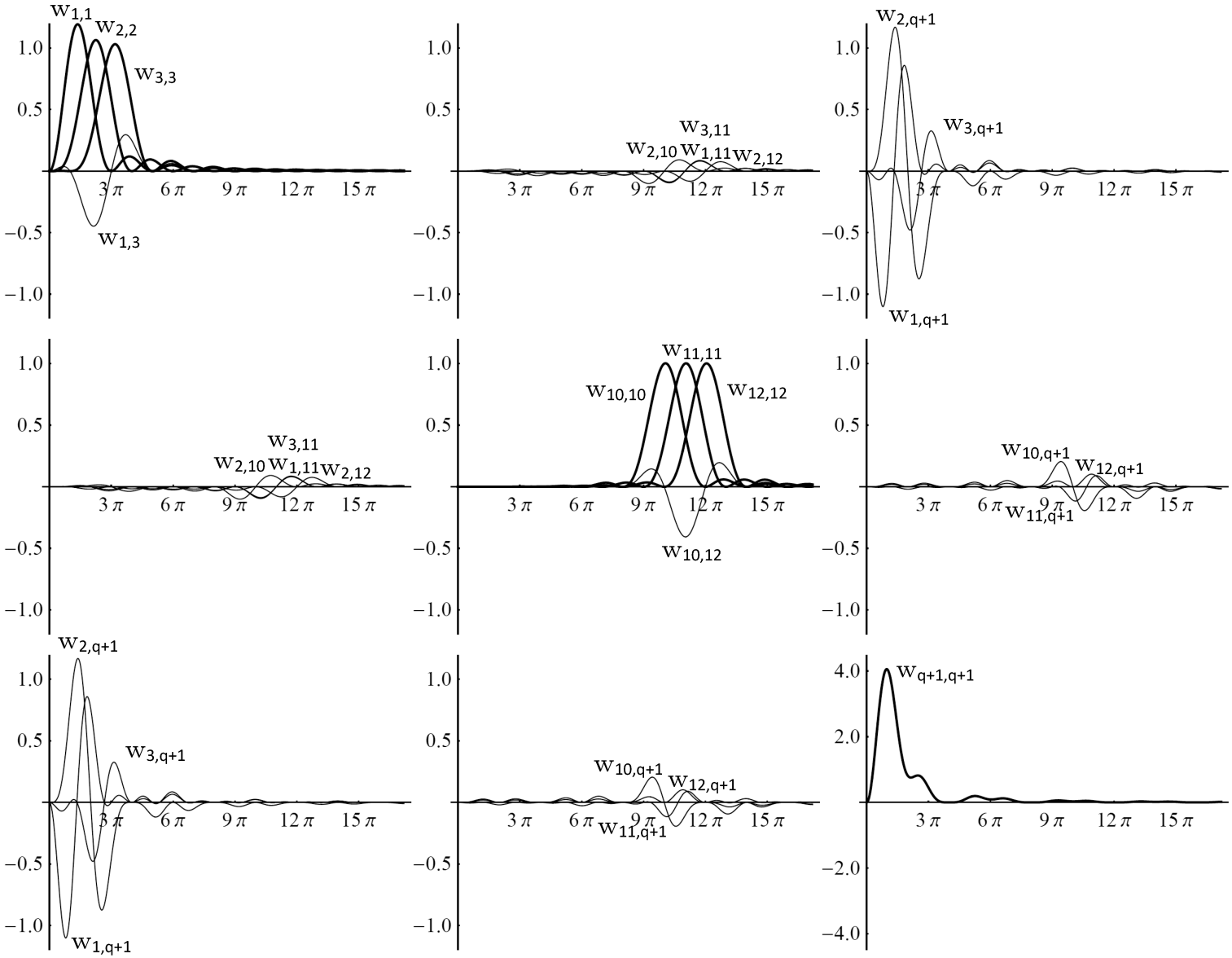
3 $\boldsymbol{\Sigma}$ as a Function of the local-to-zero Spectrum

Letting $\eta_{j,T}=T^{-1/2}\int_0^{1+r}g_j(s)x_{\lfloor sT\rfloor+1}ds$ and $\eta_{k,T}=T^{-1/2}\int_0^{1+r}g_k(s)x_{\lfloor sT\rfloor+1}ds$, Corollary 1 shows that the jkth element of $\boldsymbol{\Sigma}$ is given by

$$\Sigma_{j,k}=\int_{-\infty}^{\infty}S(\omega)\left(\int_0^{1+r}g_k(s)e^{\mathrm{i}\omega s}ds\right)\left(\int_0^{1+r}g_j(s)e^{-\mathrm{i}\omega s}ds\right). \tag{B.2}$$

Since S is an even function and g_j and g_k are real valued, (B.2) can be rewritten as $\Sigma_{j,k}=\int_0^{\infty}S(\omega)w_{jk}(\omega)d\omega$, where $w_{jk}(\omega)=\mathrm{Re}[\left(\int_0^{1+r}g_j(s)e^{-\mathrm{i}\omega s}ds\right)\left(\int_0^{1+r}g_k(s)e^{\mathrm{i}\omega s}ds\right)]$, as noted in the main text. With $g_j=\sqrt{2}\cos(\pi js)$, a calculation shows that $w_{jk}(\omega)=0$ for $1\leq j,k\leq q$ and $j+k$ odd, so that $E(X_jX_k)=0$ for all odd $j+k$, independent of the local-to-zero spectrum S. Figure B.1 plots $w_{jk}(\cdot)$ for some selected values of j and k and $r=1/2$. The figure displays the weights $w_{j,k}$ corresponding to the covariance matrix of the vector $(X_1,X_2,X_3,X_{10},X_{11},X_{12},Y)'$, organized into a symmetric matrix of 9 panels. The first panel plots the weights for the covariance matrix of $(X_1,X_2,X_3)'$, where the weights for the variances are shown in bold and the weights for the covariances are shown as thin curves. The second panel plots the weights associated with covariances between $(X_1,X_2,X_3)'$ and $(X_{10},X_{11},X_{12})'$, and so forth.

Figure B.1: Integral weights on local-to-zero spectrum for selected elements of Σ



Notes: The curves denote weights for variances (bold) and covariances (thin) of (X_1, X_2, X_3, X_{10}, X_{11}, X_{12}, Y) with $w_{j,q+1}$ corresponding to $E(X_jY)$, and $w_{q+1,q+1}$ corresponding to $E(Y^2)$. The nine panels show weights for the covariances of the three sets of variables (X_1, X_2, X_3), (X_{10}, X_{11}, X_{12}), and Y, respectively. All weights $w_{j,k}$ on covariance terms $\Sigma_{jk} = E(X_jX_k)$ with $j,k \leq q$ and $j+k$ odd are equal to zero.

4 Computation of $\boldsymbol{\Sigma}$ in *bcd*-model

One possibility to compute $\boldsymbol{\Sigma}$ is to notice that for all g_j under consideration, $\int_0^{1+r} g_j(s)e^{\mathrm{i}\omega s}ds$ can be obtained in closed form, so that one can apply one-dimensional numerical integration to (9) for $S(\omega) = (2\pi)^{-1}((c^2+\omega^2)^{-d} + b^2)$. It turns out, however, that the integrand is highly oscillatory and, for d small ($d = -0.4$), very slowly decaying.

We therefore now derive a numerically more stable expression for $\Sigma_{j,k}$ in the *bcd*-model. Note that $\boldsymbol{\Sigma}$ in (9) is linear in S, and for and $S(\omega) = (2\pi)^{-1}b^2$, $\boldsymbol{\Sigma}$ can be computed straightforwardly from (9). It thus suffices to obtain a more convenient expression for $S(\omega) = (2\pi)^{-1}(c^2+\omega^2)^{-d}$, and to add the component stemming from $b > 0$ at the end.

Now for $c > 0$ and $1/2 < d$, from (4.58) in Lindgren (2013)

$$\gamma_{c,d}(s) = (2\pi)^{-1}\int_{-\infty}^{\infty}(c^2+\omega^2)^{-d}e^{\mathrm{i}\omega s}d\omega = \frac{2^{1/2-d}\sqrt{\pi}(|s|/c)^{d-1/2}I_{d-1/2}(c|s|)}{\pi\Gamma(d)} \tag{B.3}$$

where $I_\nu(x)$ is the modified Bessel function of the second kind, so that under $b = 0$,

$$\begin{aligned}\Sigma_{j,k} &= \int_{-\infty}^{\infty}\int_0^{1+r}\int_0^{1+r} g_j(s)e^{\mathrm{i}\omega s}g_k(u)e^{-\mathrm{i}\omega u}(c^2+\omega^2)^{-d}e^{\mathrm{i}\omega s}d\omega ds du &\text{(B.4)}\\ &= \int_0^{1+r}\int_0^{1+r} g_j(s)g_k(u)\gamma_{c,d}(s-u)dsdu. &\text{(B.5)}\end{aligned}$$

We approximate (B.5) by a double sum with a step length of 0.0002. For $c = 0$, we use the implied value for $c = 0.01$.

It remains to obtain a suitable expression for $-1/2 < d < 1/2$. Suppose g is differentiable on (H_L, H_U) with derivative g'. Then by integration by parts,

$$\begin{aligned}\int_{H_L}^{H_U} g(s)e^{\mathrm{i}\omega s}ds &= \int_{H_L}^{H_U} g(s)e^{-cs}e^{(c+\mathrm{i}\omega)s}ds\\ &= g(H_U)e^{-H_U c}\frac{e^{(c+\mathrm{i}\omega)H_U}}{c+\mathrm{i}\omega} - g(H_L)\frac{1}{c+\mathrm{i}\omega} - \int_{H_L}^{H_U}(g'(s)e^{-cs} - cg(s)e^{-cs})\frac{e^{(c+\mathrm{i}\omega)s}}{c+\mathrm{i}\omega}ds\\ &= \frac{g(H_U)e^{\mathrm{i}\omega H_U} - g(H_L)e^{\mathrm{i}\omega H_L} - \int_{H_L}^{H_U}(g'(s) - cg(s))e^{\mathrm{i}\omega s}ds}{c+\mathrm{i}\omega}.\end{aligned}$$

All g functions entering $\Sigma_{j,k}$ are differentiable on $(0,1)$ and on $(1,1+r)$, so we obtain

$$\int_0^{1+r} g(s)e^{\mathrm{i}\omega s}ds = \frac{g(1+r)e^{\mathrm{i}\omega(1+r)} - (g^+(1) - g(1))e^{\mathrm{i}\omega} - g(0) - \int_0^{1+r}(g'(s) - cg(s))e^{\mathrm{i}\omega s}ds}{c+\mathrm{i}\omega} \tag{B.6}$$

where $g^+(1)$ the right limit of $g(s)$ at $s = 1$, and $g'(s)$ for $s \in \{0, 1, 1+r\}$ may be defined arbitrarily. Now applying (B.6) in (B.4) twice, and noting that $(c+\mathrm{i}\omega)(c-\mathrm{i}\omega) = c^2+\omega^2$, all integrals with respect to $d\omega$ in (B.4) are of the form $C_1\int_{-\infty}^{\infty}(c^2+\omega^2)^{-(d+1)}e^{\mathrm{i}\omega(s-C_2)}d\omega$ for C_1 and C_2 not depending on ω, which can be computed in closed form using (B.3). We obtain

$$\Sigma_{j,k} = \begin{pmatrix} g_j(1+r)\\ g_j(1) - g_j^+(1)\\ -g_j(0)\end{pmatrix}'\begin{pmatrix}\gamma_{c,d+1}(0) & \gamma_{c,d+1}(r) & \gamma_{c,d+1}(1+r)\\ \gamma_{c,d+1}(r) & \gamma_{c,d+1}(0) & \gamma_{c,d+1}(1)\\ \gamma_{c,d+1}(1+r) & \gamma_{c,d+1}(1) & \gamma_{c,d+1}(0)\end{pmatrix}\begin{pmatrix} g_k(1+r)\\ g_k(1) - g_k^+(1)\\ -g_k(0)\end{pmatrix}$$

$$
\begin{aligned}
&- \begin{pmatrix} g_j(1+r) \\ g_j(1) - g_j^+(1) \\ -g_j(0) \end{pmatrix}' \begin{pmatrix} \int_0^{1+r} (g_k'(s) - cg_k(s))\gamma_{c,d+1}(1+r-s)ds \\ \int_0^{1+r} (g_k'(s) - cg_k(s))\gamma_{c,d+1}(1-s)ds \\ \int_0^{1+r} (g_k'(s) - cg_k(s))\gamma_{c,d+1}(s)ds \end{pmatrix} \\
&- \begin{pmatrix} g_k(1+r) \\ g_k(1) - g_k^+(1) \\ -g_k(0) \end{pmatrix}' \begin{pmatrix} \int_0^{1+r} (g_j'(s) - cg_j(s))\gamma_{c,d+1}(1+r-s)ds \\ \int_0^{1+r} (g_j'(s) - cg_j(s))\gamma_{c,d+1}(1-s)ds \\ \int_0^{1+r} (g_j'(s) - cg_j(s))\gamma_{c,d+1}(s)ds \end{pmatrix} \\
&+ \int_0^{1+r} \int_0^{1+r} (g_j'(s) - cg_j(s))(g_k'(u) - cg_k(u))\gamma_{c,d+1}(s-u)dsdu.
\end{aligned}
$$

We again approximate the integrals in this expression by sums with a step length of 0.0002, and approximate $\mathbf{\Sigma}$ for $c = 0$ by the value implied by $c = 0.01$.

5 Autocovariances for the *bcd*-model

As described in the text, $S(\omega) = \frac{\sigma^2}{2\pi}((c^2 + \omega^2)^{-d} + b^2)$ is the local-to-zero spectrum of the process $x_t = e_{1t} + (bT^d)^{-1}z_t$ where $(1 - \rho_T L)^d z_t = e_{2t}$ with $\rho_T = 1 - c/T$ and e_{1t} and e_{2t} mutually uncorrelated white noise processes with common variance σ^2. The autocovariances for the process are the sum of the autocovariances for its two components. The autocovariances for the second component can be computed as follows. Suppressing the dependence of ρ on T, the spectrum for z_t is

$$R(\lambda) = \frac{\sigma^2}{2\pi}(1 + \rho^2 - 2\rho\cos(\lambda))^{-d}$$

and the kth autocovariance is therefore

$$
\begin{aligned}
\gamma_k &= \int_0^\pi \cos(k\lambda)R(\lambda)d\lambda \\
&= \frac{\sigma^2}{2\pi}\int_0^\pi \cos(k\lambda)(1 + \rho^2 - 2\rho\cos(\lambda))^{-d}d\lambda \\
&= \frac{\sigma^2}{2\pi}\left(\frac{\rho}{(1+\rho)^2}\right)^k (1+\rho)^{-2d}\left(\frac{\Gamma(d+k)}{\Gamma(d)\Gamma(1+k)}\right) {}_2F_1\left(0.5 + k, d + k; 1 + 2k; \frac{4\rho}{(1+\rho)^2}\right)
\end{aligned}
$$

where ${}_2F_1(a, b; c, z)$ is the hypergeometric function.

From Luke (1975, page 271 equation (7)):

$${}_2F_1(a, b; 2a; z) = \left(\frac{2}{1+y}\right)^{2b} {}_2F_1\left(b, b + 0.5 - a; a + 0.5; \left(\frac{1-y}{1+y}\right)^2\right), \text{ where } y = (1-z)^{1/2}.$$

Thus, the expression for γ_k simplifies:

$$
\begin{aligned}
\gamma_k &= \frac{\sigma^2}{2\pi}\left(\frac{\rho}{(1+\rho)^2}\right)^k (1+\rho)^{2(d+k)}(1+\rho)^{-2d}\left(\frac{\Gamma(d+k)}{\Gamma(d)\Gamma(1+k)}\right) {}_2F_1\left(d + k, d; 1 + k; \rho\right) \\
&= \frac{\sigma^2}{2\pi}\rho^k\left(\frac{\Gamma(d+k)}{\Gamma(d)\Gamma(1+k)}\right) {}_2F_1\left(d + k, d; 1 + k; \rho^2\right).
\end{aligned}
$$

Table B.2: Relative volume, $R_\theta(A)$, for 90% A^{MN} prediction sets, $r = 0.5$, $q = 12$

	Parameter values (θ)												
b	0	0	0	0	0	0	2	0.2	0.04	0	0	0
c	0	0	0	0	0	0	0	0	0	0.82	3.32	27.11
d	-0.40	-0.20	0.00	0.40	0.80	1.00	1	1	1	1	1	1
	(a) Sets constrained to include Bayes set											
W^{Bench}	1.85	1.88	1.80	1.54	1.27	1.15	1.78	1.38	1.22	1.89	1.19	1.40
Env	1.80	1.85	1.76	1.47	1.22	1.02	1.73	1.31	1.12	1.76	1.09	1.32
	(b) Sets unconstrained											
W^{Bench}	1.69	1.84	1.80	1.54	1.27	1.15	1.77	1.38	1.22	1.88	1.19	1.40
Env	1.57	1.79	*1.74*	1.47	1.22	1.02	1.71	1.30	1.12	1.67	1.09	1.32

Notes: The table shows the values of $R_\theta(A)$ for the value of $\theta = (b,c,d)$ shown in the first three rows. W^{Bench} denotes the benchmark weighting function described in the text, and in the rows labeled "W^{Bench}," A minimizes a weighted average of $R_\theta(A)$; in the rows labeled "Env," A is a θ-specific set that minimizes expected length at that value of θ. The non-zero values of b and c are chosen to lie between the I(0) ($b\to\infty$, $c\to\infty$) and I(1) (b=0, c=0) extremes.

6 Analysis of the Weighting Function W

As discussed in the text, construction of the A^{MN} sets requires specification of the weighting function W which determines the trade-off between expected length for various of θ. Our choice of W uses the minimized values of $V_\theta(A)$. Denote the scaled version of $V_\theta(A)$ by $R_\theta(A) = V_\theta(A)/V_\theta^{\text{known}}$, where V_θ^{known} is the expected length of the prediction set for known value of θ implied by equation (8). This relative measure of length is readily interpretable: $R_\theta(A) > 1$ represents the "regret" about not knowing θ when using the set A. In terms of $R_\theta(A)$ we use a weighting function that coincides with the Bayes prior: uniform on $d \in [-0.4, 1.0]$ and with $b = c = 0$ (so in terms of $V_\theta(A)$, the weighting function W is proportional to $1/V_\theta^{\text{known}}$). Thus, we robustify the Bayes sets in a way that minimizes regret using the Bayes prior while achieving frequentist coverage.

Table B.2 examines the expected length of the $A^{MN} = A^{MN}_{(b,c,d)}$ sets and answers questions regarding the choice of W and Γ. Because A^{MN} minimizes a weighted average of the values of the expected (normalized) length $R_\theta(A)$, it is impossible to choose an alternative weighting function with uniformly lower values of $R_\theta(A)$. However, it is possible to reduce $R_\theta(A^{MN})$ for any particular value of θ. Thus, Table B.2 compares the value $R_\theta(A^{MN})$ for the benchmark choice of W to the value obtained by putting all W mass on θ, for selected values of θ. Said differently, Table B.2 compares $R_\theta(A^{MN})$ to the envelope obtained by minimizing $R_\theta(A)$ over all equivariant sets A that have uniform coverage in the *bcd*-model. Values of $R_\theta(A)$ are shown as a function of d in the $I(d)$ model, and for selected values in the b and c models, respectively.

The table shows results for 90% prediction sets; conclusions for 67% prediction sets are similar. Panel (a) considers sets that are also restricted to contain the Bayes prediction set, as in (13). For the benchmark choice of W, $R_\theta(A) = R_\theta(A^{MN})$ ranges from 1.15 (so that $V_\theta(A^{MN})$ is 15% larger

Table B.3: Average length of 90% A^{MN} prediction sets, $r = 0.5$

	$q = 12$		$q = 6$	$q = 24$	$q = 48$
θ unknown	1.57		1.69	1.45	1.38
θ known	1.00		1.08	0.97	0.95

Notes: Entries are weighted average regret (relative to knowing θ in the q=12 case) of A^{MN} prediction sets (θ unknown) and minimum length sets A_θ^{known} for θ known.

than the attainable expected length with θ known) when $b = c = 0$ and $d = 1.0$, to 1.89 when $b = 0$, $c = 27$, and $d = 1.0$. These numbers demonstrate that parameter uncertainty is a significant component of A^{MN}. The envelope numbers are smaller, of course, but are typically within 10% of the A^{MN} values even for θ with non-zero b or c (where the benchmark weighting function puts zero mass).

Panel (b) of Table B.2 takes up the question of the cost of imposing the Bayes-superset constraint (13). The two rows in panel (b) are computed exactly as those in panel (a), except that the Bayes-superset constraint is no longer enforced. Evidently, for most values of θ, the Bayes-superset constraint has a negligible impact of the expected length of the prediction set. What is more, the envelope numbers in panel (b) are typically not much smaller than those for A^{MN} in panel (a), which means that the substantial regret of not knowing θ is mostly driven by the coverage constraint, and is not an artifact of our benchmark choices for Γ or W. (And, as a corollary, it also implies that alternative choices for Γ for A^{MN} can not lead to substantially shorter sets on average).

7 Choice of q

As discussed in Section 4.3, the sample cosine transforms contain information on both the scale and shape of the local-to-zero spectrum. Table B.3 quantifies the combined scale and shape effects by reporting the value of the objective $\int V_\theta(A)dW(\theta)$ in the program (10) for $q \in \{6, 12, 24, 48\}$. To be able to make meaningful comparisons across different values of q, we choose W as the benchmark weighting function from the $q = 12$ case for all values of q, normalized to integrate to one (so that the values in Table 3 are the weighted average regret relative to knowing θ in the $q = 12$ case). In this θ unknown case, there is an 8% decrease in average length as q increases from $q = 12$ to $q = 24$ and a further reduction of 5% for $q = 48$. As a point of reference, the table also reports the same average length for θ known.[16] The reductions in lengths for $q > 12$ are only marginally larger than the pure scale effect in the $I(0)$ model, corroborating the intuition above about the relative unimportance of the spectral shape outside the $[-12\pi/T, 12\pi/T]$ interval for long-run forecasting.

[16] These are for comparison only; the parameters of the *bcd*-model cannot be consistently estimated even under $q \to \infty$ asymptotics.

8 Computation of Prediction Sets

8.1 Numerical Determination of $C_\theta(A)$ and $V_W(A)$

The algorithm for determining A_{MN} below requires repeated evaluation of $V_\theta(A)$ and $C_\theta(A)$, with A some set valued function of $\mathbf{x}^s$. We employ an importance sampling Monte Carlo scheme: Write f_θ for the density of $\mathbf{Z}^s = (\mathbf{X}^s, Y^s)$ under θ, i.e. $f_\theta(\mathbf{z}^s) = f_{(\mathbf{X}^s,Y^s)|\theta}(\mathbf{x}^s, y^s)$, and let f_p be a proposal density for $\mathbf{Z}^s$ such that f_θ is absolutely continuous with respect to f_p for all θ. Then

$$C_\theta(A) = E_p[\frac{f_\theta(\mathbf{Z}^s)}{f_p(\mathbf{Z}^s)}\mathbf{1}[Y^s \in A(\mathbf{X}^s)]]$$

with E_p denoting integration with respect to f_p. Furthermore, from

$$\begin{aligned} V_\theta(A) &= E_\theta\left[g_\theta(\mathbf{X}^s)\,\mathrm{vol}(A(\mathbf{X}^s))\right] \\ &= E_\theta\left[g_\theta(\mathbf{X}^s)\int \mathbf{1}[y^s \in A(\mathbf{X}^s)]dy^s\right] \\ &= \int\int g_\theta(\mathbf{x}^s) f_{\mathbf{X}^s|\theta}(\mathbf{x}^s)\mathbf{1}[y^s \in A(\mathbf{x}^s)]d\mathbf{x}^s dy^s \end{aligned}$$

we obtain

$$V_\theta(A) = E_p[\frac{g_\theta(\mathbf{X}^s) f_{\mathbf{X}^s|\theta}(\mathbf{X}^s)}{f_p(\mathbf{Z}^s)}\mathbf{1}[Y^s \in A(\mathbf{X}^s)]].$$

With N i.i.d. draws $\mathbf{Z}_1^s, \ldots, \mathbf{Z}_m^s$ from f_p, we thus obtain the importance sampling approximations

$$\hat{C}_\theta(A) = N^{-1}\sum_{l=1}^{N}\frac{f_\theta(\mathbf{Z}_l^s)}{f_p(\mathbf{Z}_l^s)}\mathbf{1}[Y_l^s \in A(\mathbf{X}_l^s)] \tag{B.7}$$

$$\hat{V}_\theta(A) = N^{-1}\sum_{l=1}^{N}\frac{g_\theta(\mathbf{X}_l^s) f_{\mathbf{X}^s|\theta}(\mathbf{X}_l^s)}{f_p(\mathbf{Z}_l^s)}\mathbf{1}[Y_l^s \in A(\mathbf{X}_l^s)]. \tag{B.8}$$

The expression (B.8) is numerically advantageous, as it does not require any explicit numerical determination of the length of a set $A(\mathbf{x}^s)$.

8.2 Numerical Determination of an ALFD

Our algorithm is analogous to what is suggested in Elliott, Müller, and Watson (2015) (EMW). It consists of two main parts: (I) determination of a candidate ϵ-ALFD Λ^*; (II) numerical check of $\inf_{\theta\in\Theta} P_\theta(Y^s \in A_{\Lambda^*,\mathrm{cv}^{*\epsilon}}(\mathbf{X}^s)) \geq 1-\alpha$. Both parts are based on a discrete grid of values for θ, with the grid Θ_{II} for part (II) finer and with a wider range than the grid Θ_I for part (I). In particular, with $\Delta_I = \{-0.4, -0.2, \ldots, 1.0\}$, $\Delta_{II} = \{-0.4, -0.3, \ldots, 1.0\}$, $\tilde{B}_I = \{0\}\cup\{e^{-5}, e^{-4}, \ldots, e^5\}$, $\tilde{B}_{II} = \{0\}\cup\{e^{-5.5}, e^{-5.0}, \ldots, e^{7.5}\}$, $C_I = \{0\}\cup\{e^{-3.0}, e^{-2.3}, \ldots, e^{4.0}\}$, $C_{II} = \{0\}\cup\{e^{-3.35}, e^{-3.0}, \ldots, e^{7.5}\}$, $\Theta_I = \{(b,0,d)' : d \in \Delta_I, (8\pi)^{2d}b^2 \in \tilde{B}_I\}\cup\{(0,c,d)' : d \in \Delta_I, c \in C_I\}\cup\{(b,c,1.0)' : c \in C_I, ((8\pi)^2 + c^2)^d b^2 \in \tilde{B}_I\}$ and $\Theta_{II} = \{(b,c,d)' : c \in C_{II}, d \in \Delta_{II}, ((8\pi)^2 + a^2)^d b^2 \in \tilde{B}_{II}\}$. Also let $\Theta_\Gamma = \{\theta : \theta = (0,0,d), d \in \{-0.4, -0.2, \ldots, 1.0\}\}$ denote the support of the prior Γ.

The range of values for b and c in grid Θ_{II} are chosen to approximately cover all possible values of $\mathbf{\Sigma}$ in the *bcd*-model (as $b \to \infty$ or $c \to \infty$, $\mathbf{\Sigma}$ converges to its value in the I(0) model, for any value of $d \in \Delta_{II}$). Preliminary results suggested that the ALFD puts no mass on θ with $b > 0$, $c > 0$ and $d < 1$, motivating our smaller choice for Θ_I.

With Θ_I the support of candidate ALFDs Λ, and W putting weight w_θ on $\theta \in \Theta_\Gamma$, (A.6) becomes

$$A_{\Lambda,\mathrm{cv}}(\mathbf{x}^s) = \left\{ y^s : \frac{\sum_{\theta\in\Theta_I} \lambda_\theta f_{(Y^s,\mathbf{X}^s)|\theta}(y^s,\mathbf{x}^s)}{\sum_{\theta\in\Theta_\Gamma} w_\theta g_\theta(\mathbf{x}^s) f_{\mathbf{X}^s|\theta}(\mathbf{x}^s)} > \mathrm{cv} \right\}$$

where $\lambda_\theta \geq 0$ and $\sum_{\theta\in\Theta_I} \lambda_\theta = 1$. Note that $A_{\Lambda,\mathrm{cv}}$ is fully determined by the ratios $\kappa_\theta = \lambda_\theta/\,\mathrm{cv}$, $\theta \in \Theta_I$, so introduce the notation $A_{\Lambda,\mathrm{cv}} = A_\kappa$.

Part (I) of the algorithm seeks to determine appropriate values of λ_θ, $\theta \in \Theta_I$, for an ϵ-ALFD. The following steps are for the A^{MN} set that also enforces the Bayes superset constraint (13).

1. Generate i.i.d. draws $\mathbf{Z}_l^s$, $l = 1,\ldots,N$ from $f_p(\mathbf{z}^s) = |\Theta_I|^{-1}\sum_{\theta\in\Theta_I} f_\theta(\mathbf{z}^s)$ (that is, the proposal density f_p is the equal probability mixture of f_θ with θ drawn uniformly from Θ_I).

2. Compute and store $f_\theta(\mathbf{Z}_l^s)$ and $f_p(\mathbf{Z}_l^s)$ for $\theta \in \Theta_I$, and $g_\theta(\mathbf{X}_l^s) f_{\mathbf{X}^s|\theta}(\mathbf{X}_l^s)$ for $\theta \in \Theta_\Gamma$, $l = 1,\ldots,N$.

3. Compute $V_\theta^{\mathrm{known}} = \hat{V}_\theta(A_\theta^{\mathrm{known}})$, $\theta \in \Theta_\Gamma$ where $A_\theta^{\mathrm{known}}$ is the prediction set for known θ. To be specific, $A_\theta^{\mathrm{known}} = [\mu_\theta(\mathbf{x}^s) - t^q_{(1-\alpha/2)}\sigma_\theta(\mathbf{x}^s), \mu_\theta(\mathbf{x}^s) + t^q_{(1-\alpha/2)}\sigma_\theta(\mathbf{x}^s)]$ with $\mu_\theta(\mathbf{x}^s) = \mathbf{\Sigma}_{YX}\mathbf{\Sigma}_{XX}^{-1}\mathbf{x}^s$ and $\sigma_\theta^2(\mathbf{x}^s) = (\Sigma_{YY} - \mathbf{\Sigma}_{YX}\mathbf{\Sigma}_{XX}^{-1}\mathbf{\Sigma}_{XY})\mathbf{x}^{s\prime}\mathbf{\Sigma}_{XX}^{-1}\mathbf{x}^s/q$, with $\mathbf{\Sigma}_{XX}$, $\mathbf{\Sigma}_{YX}$, $\mathbf{\Sigma}_{XY}$ and Σ_{YY} as implied by θ. Set $w_\theta = (V_\theta^{\mathrm{known}})^{-1}/\sum_{t\in\Gamma_\theta}(V_t^{\mathrm{known}})^{-1}$, $\theta \in \Gamma_\theta$.

4. For each $\mathbf{Z}_l^s$, determine $\mathbf{1}[Y_l^s \in A^{\mathrm{Bayes}}(\mathbf{X}_l^s)] = \mathbf{1}[F^{\mathrm{Bayes}}(Y_l^s|\mathbf{X}_l^s) \in (\alpha/2, 1-\alpha/2)]$, where F^{Bayes} is the c.d.f. of the Bayes predictive distribution, that is

$$F^{\mathrm{Bayes}}(Y^s|\mathbf{X}^s) = \frac{\sum_{\theta\in\Theta_\Gamma} \gamma_\theta f_{\mathbf{X}^s|\theta}(\mathbf{X}^s) F_t^q\left(\frac{Y^s - \mu_\theta(\mathbf{X}^s)}{\sigma_\theta(\mathbf{X}^s)}\right)}{\sum_{\theta\in\Theta_\Gamma} \gamma_\theta f_{\mathbf{X}^s|\theta}(\mathbf{X}^s)}$$

   with F_t^q the c.d.f. of a student-t distribution with q degrees of freedom and $\gamma_\theta = 1/|\Theta_\Gamma|$.

5. Set $\kappa^{(0)} = (1,\ldots,1) \in \mathbb{R}^{|\Theta_I|}$, $\omega^{(0)} = (4,\ldots,4) \in \mathbb{R}^{|\Theta_I|}$, $\mathrm{RP}^{(0)} = (\alpha,\ldots,\alpha) \in \mathbb{R}^{|\Theta_I|}$

6. Iterate over $i = 1,\ldots,500$

   (a) Compute $\mathrm{RP}_\theta^{(i)} = 1 - \hat{C}_\theta(A_{\kappa^{(i)}} \cup A^{\mathrm{Bayes}})$, $\theta \in \Theta_I$.

   (b) Set $\kappa_\theta^{(i+1)} = \kappa_\theta^{(i)} \exp(\omega_\theta^{(i)}(\mathrm{RP}_\theta^{(i)} - \alpha))$, $\theta \in \Theta_I$.

   (c) If $i \geq 400$, set $\omega_\theta^{(i+1)} = 0.1$. Otherwise, set $\omega_\theta^{(i+1)} = 1.03\omega_\theta^{(i)}$ if $(\mathrm{RP}_\theta^{(i)} - \alpha)(\mathrm{RP}_\theta^{(i)} - \alpha) > 0$ and $\omega_\theta^{(i+1)} = 0.5\omega_\theta^{(i)}$ otherwise, $\theta \in \Theta_I$.

7. Set $\lambda_\theta^* = \kappa_\theta^{(500)}/\sum_{t\in\Theta_I}\kappa_t^{(500)}$, $\theta \in \Theta_I$.

As in EMW, the simple idea underlying Step 6.b is a fixed-point iteration that decreases κ_θ under overcoverage, and increases it otherwise, with the change (in logarithms) proportional to the degree of over- or undercoverage. The idea of Step 6.c is to half the step length if the sign of the coverage violation switched in the last iteration, and to gradually increase it otherwise. In the last 100 of the 500 iterations, a small step length is enforced regardless.

One would expect λ_θ^* of Step 7 to be a rough approximation to the least favorable distribution with the parameter space restricted to Θ_I. With Θ_I a sufficiently fine grid, this in turn is a candidate for an ϵ-ALFD Λ^* in the sense of Definition 2. Part (II) of the algorithm checks whether Λ^* constructed from λ_θ^* does indeed satisfy the properties of an ϵ-ALFD Λ^*.

1. Compute the number cv* such that $\sum_{\theta\in\Theta_\Gamma} w_\theta \hat{C}_\theta(A_{\Lambda^*,\text{cv}^*} \cup A^{\text{Bayes}}) = 1-\alpha$.
2. Compute the number cv$^{*\epsilon}$ such that $\sum_{\theta\in\Theta_\Gamma} w_\theta \hat{V}_\theta(A_{\Lambda^*,\text{cv}^{*\epsilon}} \cup A^{\text{Bayes}}) = \epsilon + \sum_{\theta\in\Theta_\Gamma} w_\theta \hat{V}_\theta(A_{\Lambda^*,\text{cv}^*} \cup A^{\text{Bayes}})$.
3. Check whether $\inf_{\theta\in\Theta_{II}} \hat{C}_\theta(A_{\Lambda^*,\text{cv}^{*\epsilon}} \cup A^{\text{Bayes}}) \geq 1-\alpha$.

We set $\epsilon = 0.01$ throughout, so that in terms of the W-average expected length, the set $A^{MN} = A_{\Lambda^*,\text{cv}^{*\epsilon}} \cup A^{\text{Bayes}}$ is at most 1% longer than any set satisfying (11)-(13). We use $N = 250,000$, leading to Monte Carlo errors of $\hat{C}_\theta$ of approximately 0.1-0.3%. For a given value of q, α and r, these computations take approximately one minute on a modern PC using Fortran. We compute A^{MN} sets for $q \in \{6, 12, 24, 48\}$ and $r \in \{0.075, 0.1, 0.15\} \cup \{0.2, 0.3, \ldots, 1.0\} \cup \{1.2, 1.4, 1.6, 1.8\}$. In the empirical work, prediction intervals for other values of r are computed via linear interpolation.

Figure B.2 displays the approximate least favorable distributions (ALFDs) that underlie the $A^{MN}_{(b,c,d)}$ sets. The ALFDs are displayed in terms of the local-to-zero spectral shapes from Figure 3 in the text, with thicker curves representing more mass in the ALFD. The ALFD is non-degenerate and has most of its mass on spectra that are relatively flat for larger ω, but with a pronounced pole at zero (these spectra arise, for instance, in the local-level with moderate b). Intuitively, in the local-level model, the strong mean reversion of the $I(0)$ component masks the pronounced long-run uncertainty, making it relatively hardest to control coverage.

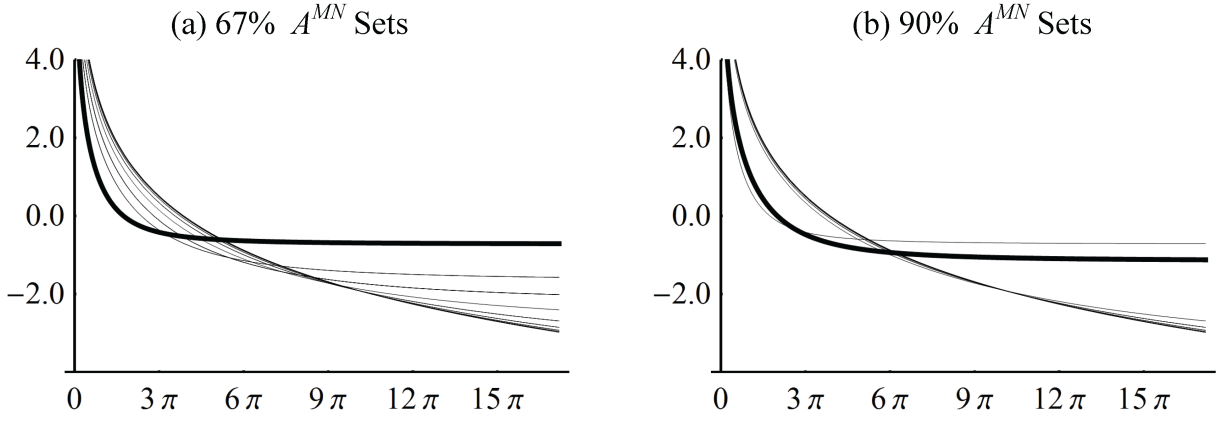
9 Parameter values in the Monte Carlo Simulations

Models 1-4 are calibrated to quarterly data and set $T = 260$ quarters. Model 5 is calibrated to monthly data and sets $T = 780$ months.

Model 1: $x_t = \mu_t + u_t$ with $u_t \sim iid\mathcal{N}(0, 3.7^2)$, $\Delta\mu_t = s_t\delta_t$, where $s_t \sim iid$Bernoulli(p), and $\delta_t \sim iid(-\delta$ with probability $= 0.5, +\delta$ with probability $= 0.5)$. The experiments use $(p, \delta) = (0, 0)$, $(1/40, 0.125)$, $(1/40, 0.375)$, $(1/260, 0.250)$, and $(1/260, 0.750)$.

Model 2: $x_t = \mu_t + u_t$ with $(1 - 0.98L)u_t = \varepsilon_t \sim iid\mathcal{N}(0, 0.46^2)$, $\Delta\mu_t = s_t\delta_t$, where $s_t \sim iid$Bernoulli(p), and $\delta_t \sim iid(+\delta$ with probability $= 0.5, +\delta$ with probability $= 0.5)$. The experiments use $(p, \delta) = (0, 0)$, $(1/40, 0.5)$, $(1/40, 1.0)$, $(1/260, 1.0)$, and $(1/260, 2.0)$.

Figure B.2: ALFDs in terms of implied local-to-zero spectra



Notes: The panels represent the logarithm of the local-to-zero spectra that receive more than 1% mass in the approximate least favorable distribution underlying the A^{MN} predictive sets for r=0.5 and q=12. Line widths are proportional to mass in the ALFD. Same scale normalization as in Figure 3.

Model 3: $x_t = \sigma_t u_t$ with $u_t \sim iid\mathcal{N}(0,1)$, $\sigma_0 = 1$, $\Delta \ln(\sigma_t) = s_t \delta_t$, where $s_t \sim iid$ Bernoulli(p), and $\delta_t \sim iid(+\delta$ with probability $= 0.5, +\delta$ with probability $= 0.5)$. The experiments use $(p, \delta) = (0,0)$, $(1/40, 0.0625)$, $(1/40, 0.1875)$, $(1/260, 0.125)$, and $(1/260, 0.375)$.

Model 4: $x_t = u_{1t} + \sigma_t u_{2t}$ with $u_{1t} \sim iid\mathcal{N}(0,1)$, $(1-L)u_{2t} = \varepsilon_t \sim iid\mathcal{N}(0, 0.707^2)$, $\sigma_0 = 1$, $\Delta \ln(\sigma_t) = s_t \delta_t$, where $s_t \sim iid$ Bernoulli(p), and $\delta_t \sim iid(+\delta$ with probability $= 0.5, +\delta$ with probability $= 0.5)$. The experiments use $(p, \delta) = (0,0)$, $(1/40, 0.407)$, $(1/40, 0.806)$, $(1/260, 0.814)$, and $(1/260, 1.612)$.

Model 5: Within regime k, $\Delta x_t = \alpha_k \phi_k - \phi_k x_{t-1} + \varepsilon_t$, with $\varepsilon_t \sim iid\mathcal{N}(0, \sigma_k^2)$, where $\alpha_k \sim \max(0, \mathcal{N}(\mu_\alpha, \sigma_\alpha^2))$, $\phi_k \sim \mathcal{N}(\mu_\phi, \sigma_\phi^2)$ truncated on $(0,1)$, and $\sigma_k^{-1} \sim$ Gamma(v, d). Regime k ends with probability $1 - p_k$, where $p_k \sim$ Beta(a, b). Values of $\mu_a, \sigma_\alpha, \mu_\phi, \sigma_\phi, v$, and d were taken from Pesaran, Pettenuzzo, and Timmermann (2006), Table 5: $\mu_a = 1.96, \sigma_\alpha^2 = 125.3, \mu_\phi = 0.084, \sigma_\phi^2 = .179, v = .884$, and $d = .050$. Values of a, b are not given in Pesaran, Pettenuzzo, and Timmermann (2006), and we take them from Pettenuzzo and Timmermann (2011), Table 4: $a = 34.77$, and $b = 0.81$. Sample sizes of $2T$ are generated and the first T observations are discarded.

Additional References

Davidson, J. (1994): *Stochastic Limit Theory.* Oxford University Press, New York.

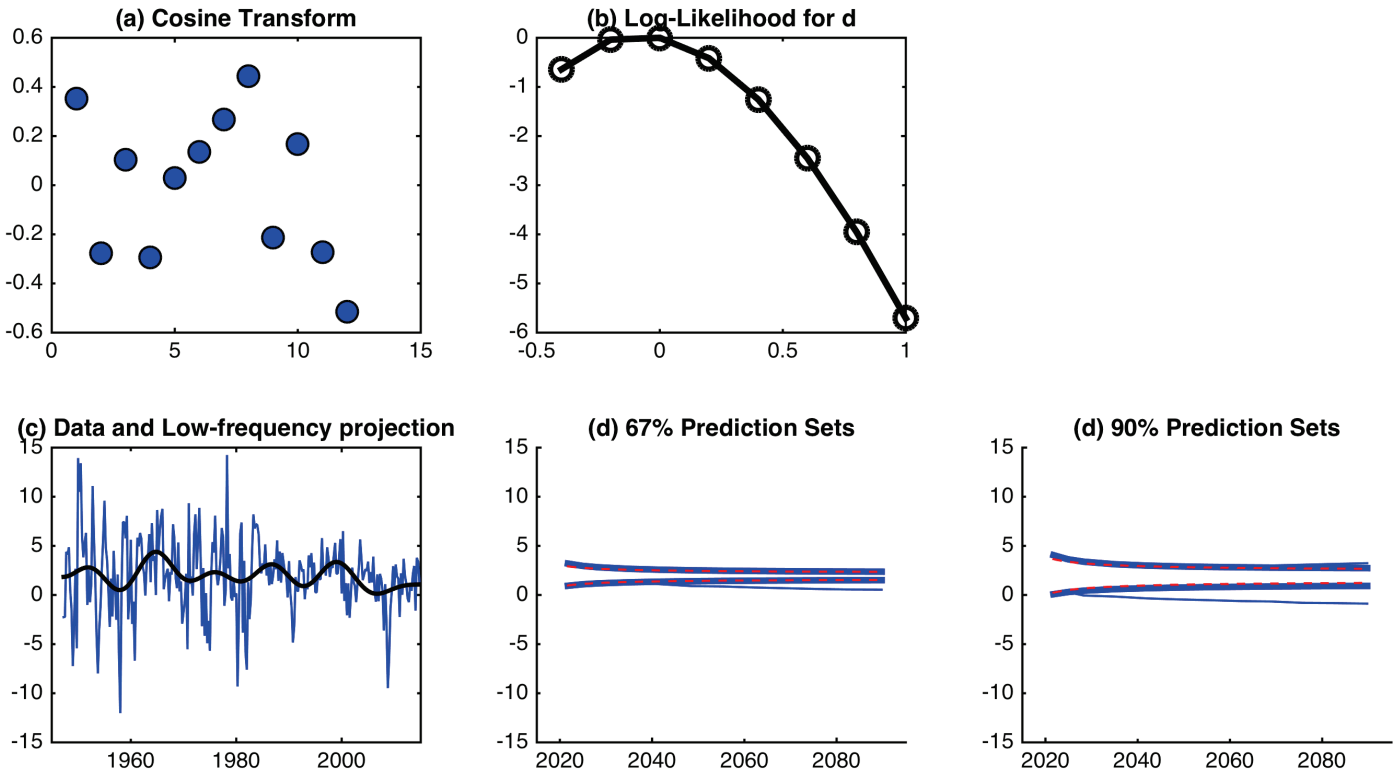
Fernald, J. (2014): "A Quarterly, Utilization-Adjusted Series on Total Factor Productivity," *FRBSF Working Paper 2012-9.*

Luke, Y. (1975): *Mathematical Functions and Their Approximations.* Academic Press, New York.

Pettenuzzo, D., and A. Timmermann (2011): "Predictability of stock returns and asset allocation under structural breaks," *Journal of Econometrics*, 164(1), 60–78.

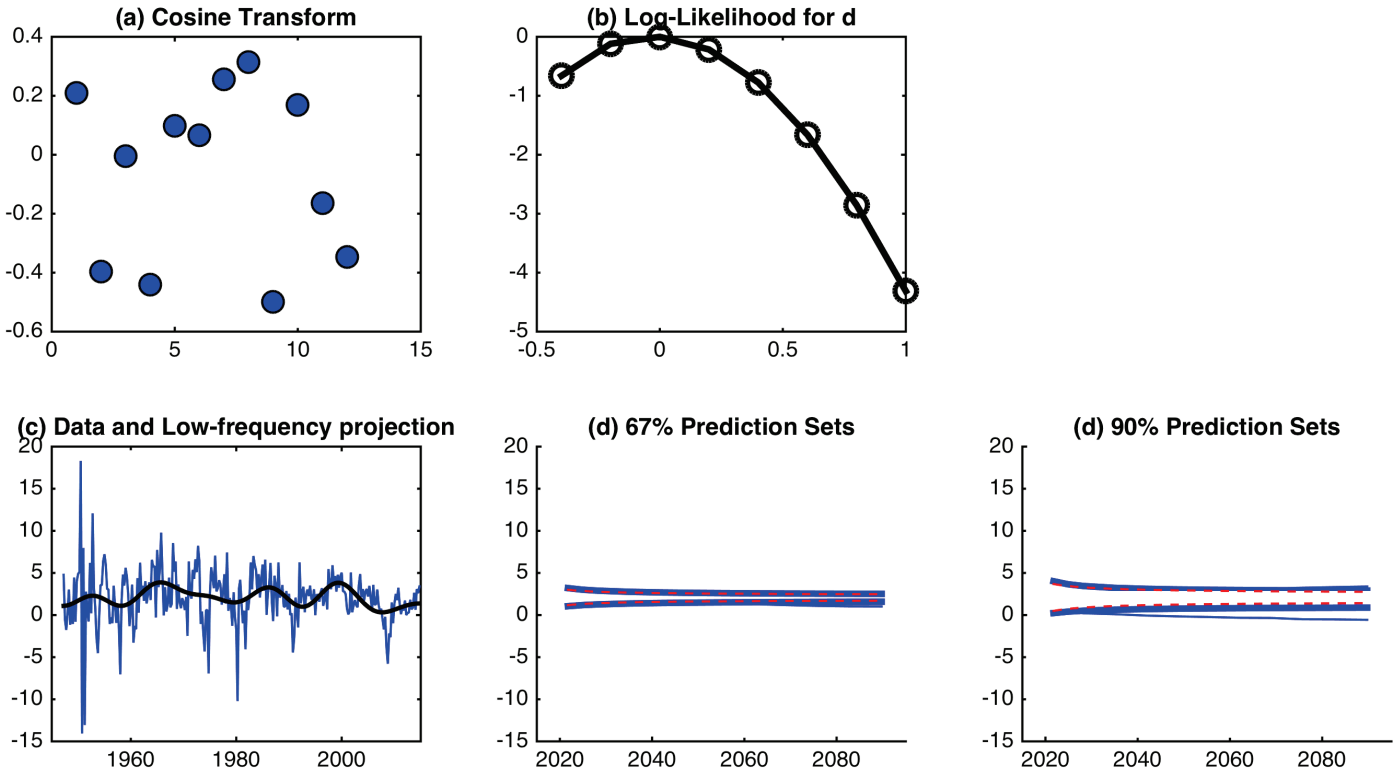
Additional Figures

C.1 Growth rate of real per-capita GDP
Sample period: 1947:Q2 – 2014:Q4, $\bar{X}_{1:T} = 1.94$, $s_{LR} = 4.82$



Notes: The cosine transformations shown in (ii) are the standardized values, $X^{s}_{T,1q}$. The low-frequency $I(d)$ likelihood is computed using $X^{s}_{T,1q}$ and its asymptotic distribution given in the text; values are relative to the $I(0)$ model. The low-frequency components in (iv) are the projection of the series onto $cos[(t–0.5)\pi j/T]$ for $j = 0, \ldots, 12$. The predictions sets in panels (v) and (vi) are A^{Bayes} (thick line), A^{MN} (thin line), and $A^{I(0)}$ (dashed line), and are computed separately for each horizon, so coverage levels hold pointwise, not jointly across horizons.

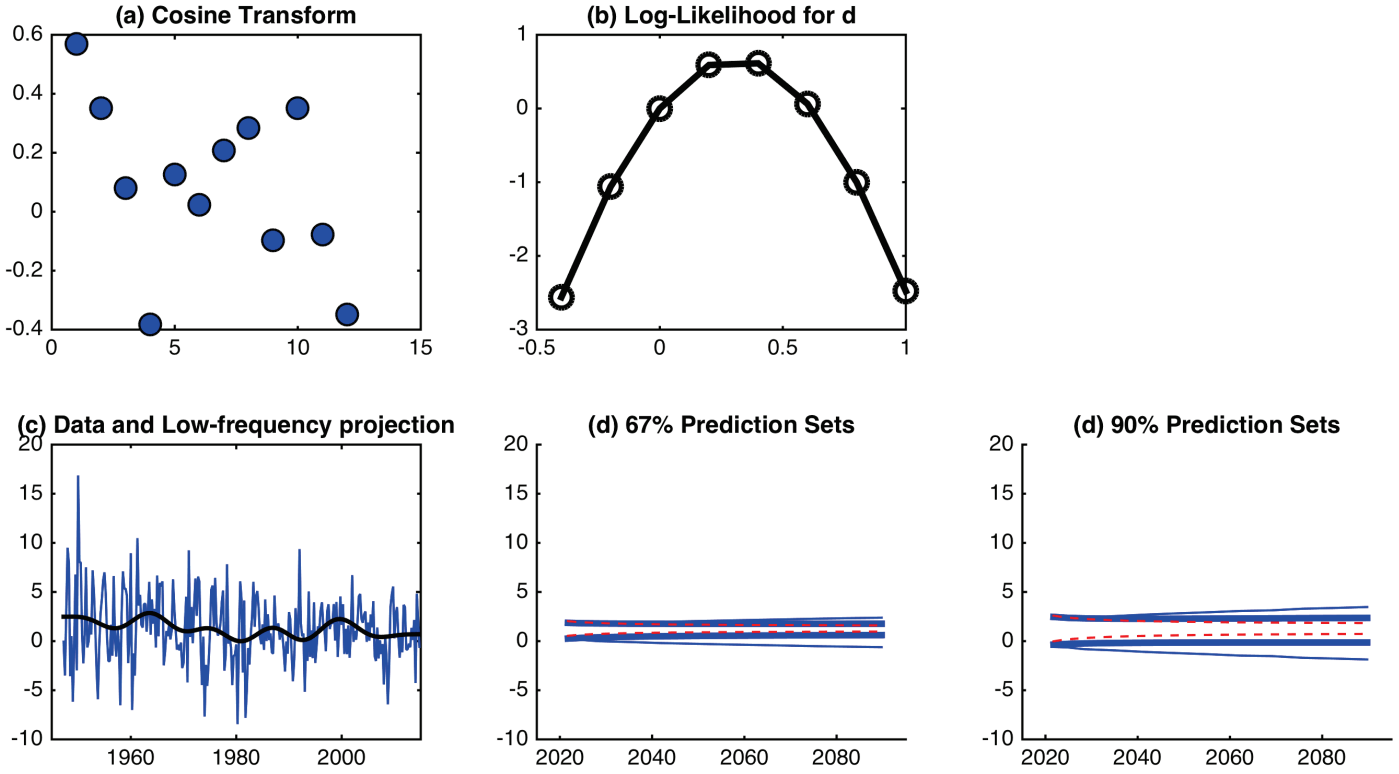
C.2 Growth rate of real per-capita consumption expenditures
Sample period: 1947:Q2 – 2014:Q4, $\bar{x}_{1:T}$ = 2.09, s_{LR} = 4.53



See notes to C.1.

C.3 Growth rate of total factor productivity

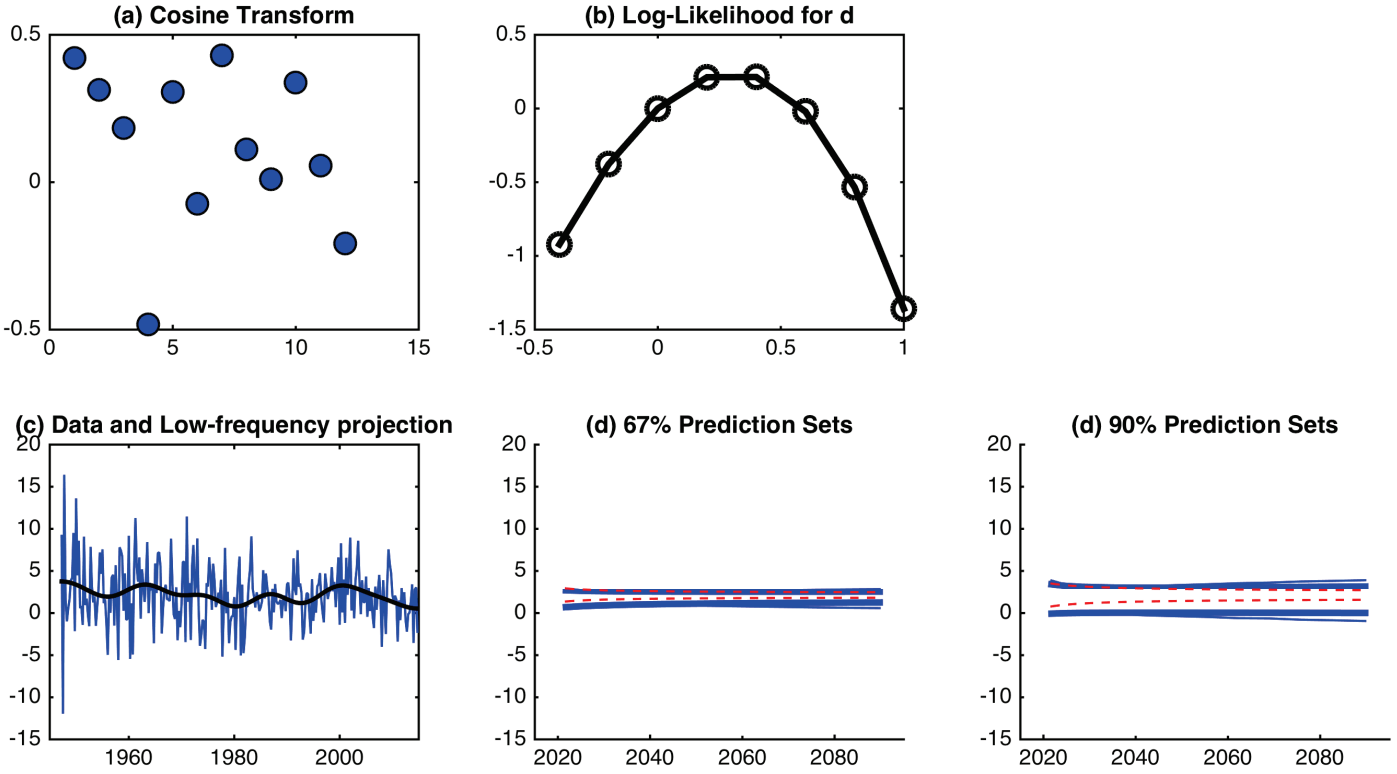
Sample period: 1947:Q2 – 2014:Q4, $\bar{x}_{1:T} = 1.29$, $s_{LR} = 3.70$



See notes to C.1.

C.4 Growth rate of labor productivity

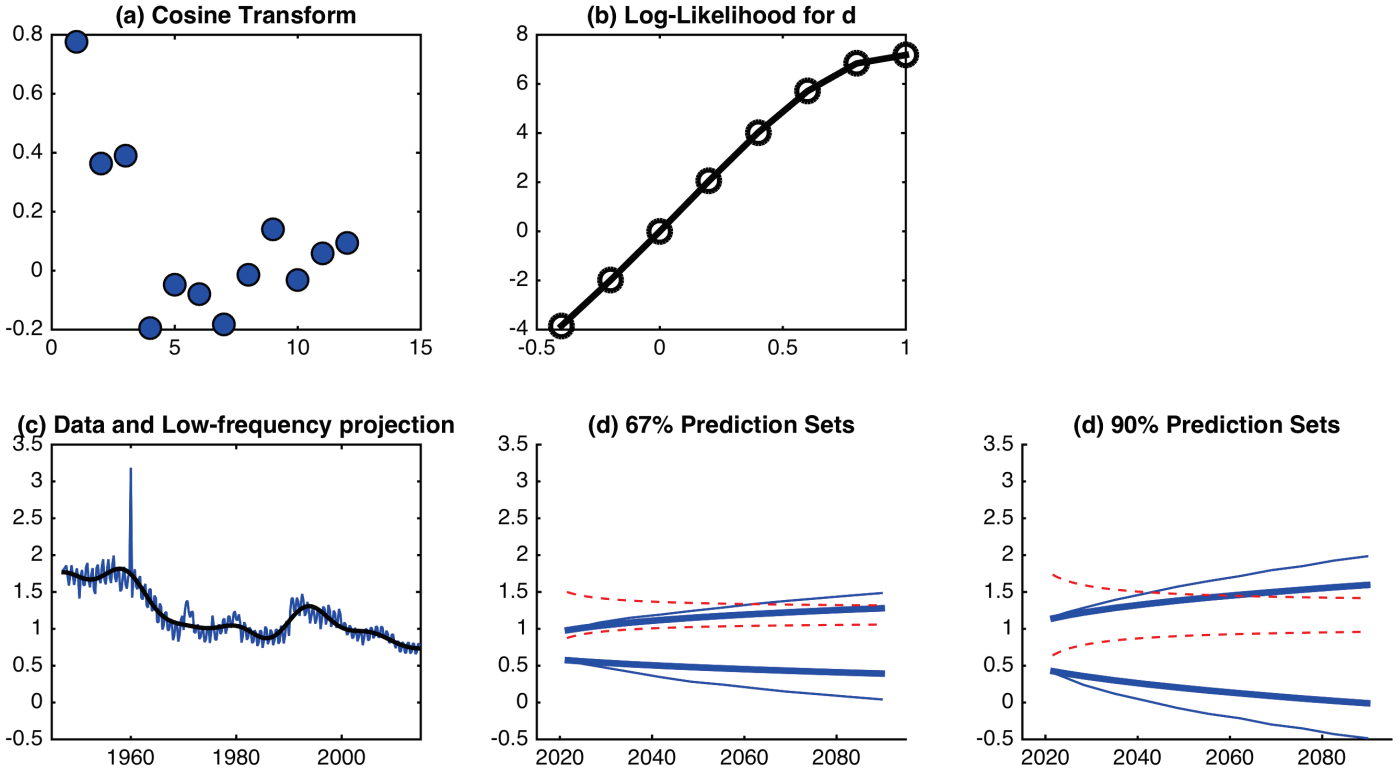
Sample period: 1947:Q2 – 2014:Q4, $\bar{x}_{1:T} = 2.15$, $s_{LR} = 3.84$



See notes to C.1.

C.5 Growth rate of population

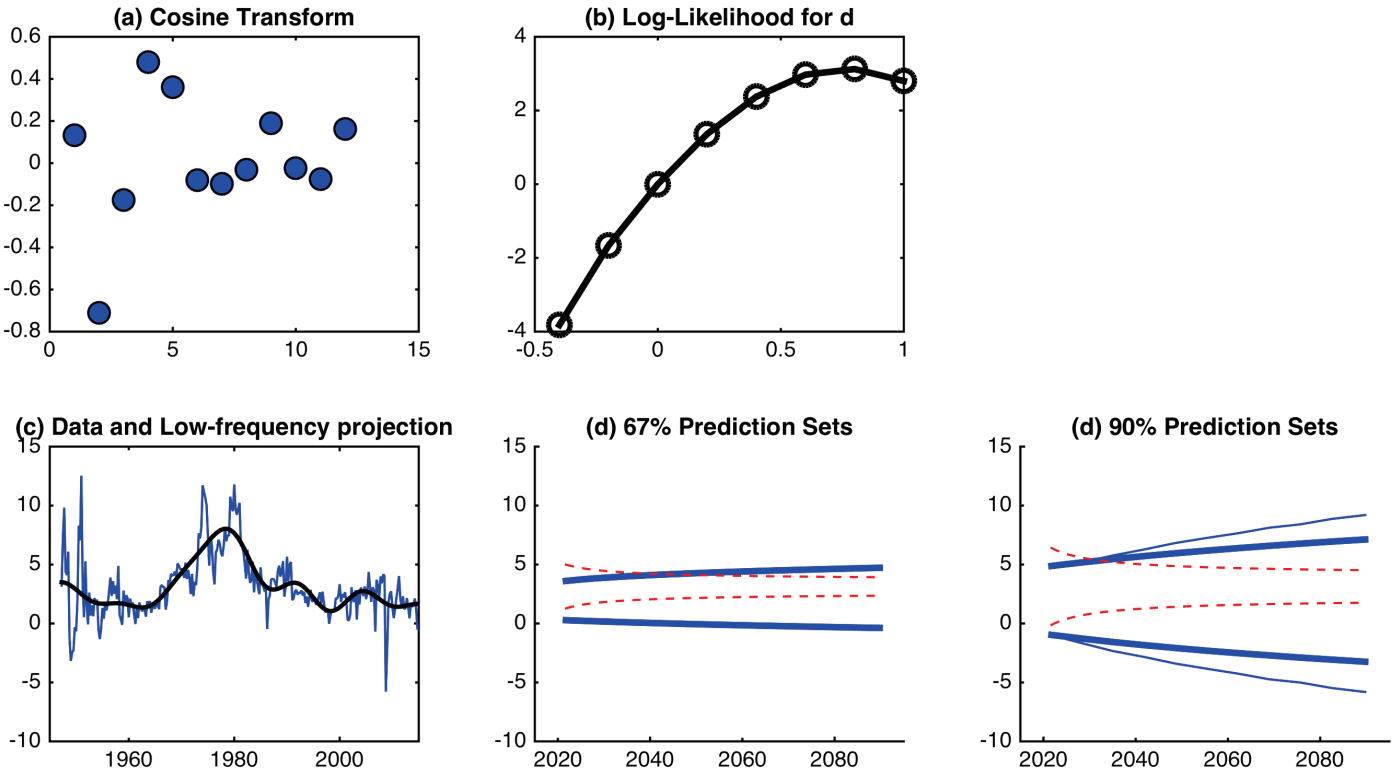
Sample period: 1947:Q2 – 2014:Q4, $\bar{x}_{1:T} = 1.19$, $s_{LR} = 1.53$



See notes to C.1.

C.6 Inflation (PCE)

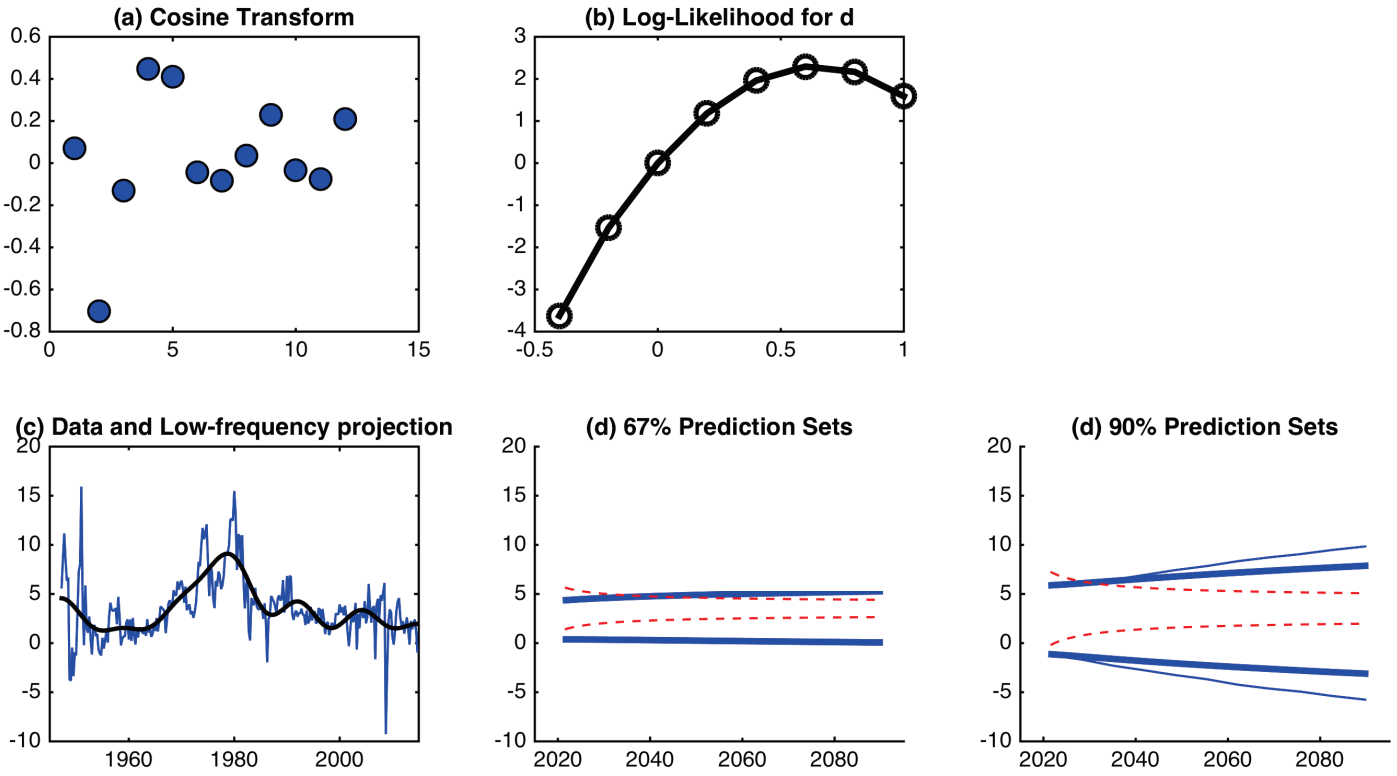
Sample period: 1947:Q2 – 2014:Q4, $\bar{x}_{1:T} = 3.14$, $s_{LR} = 9.22$



See notes to C.1.

C.7 Inflation (CPI)

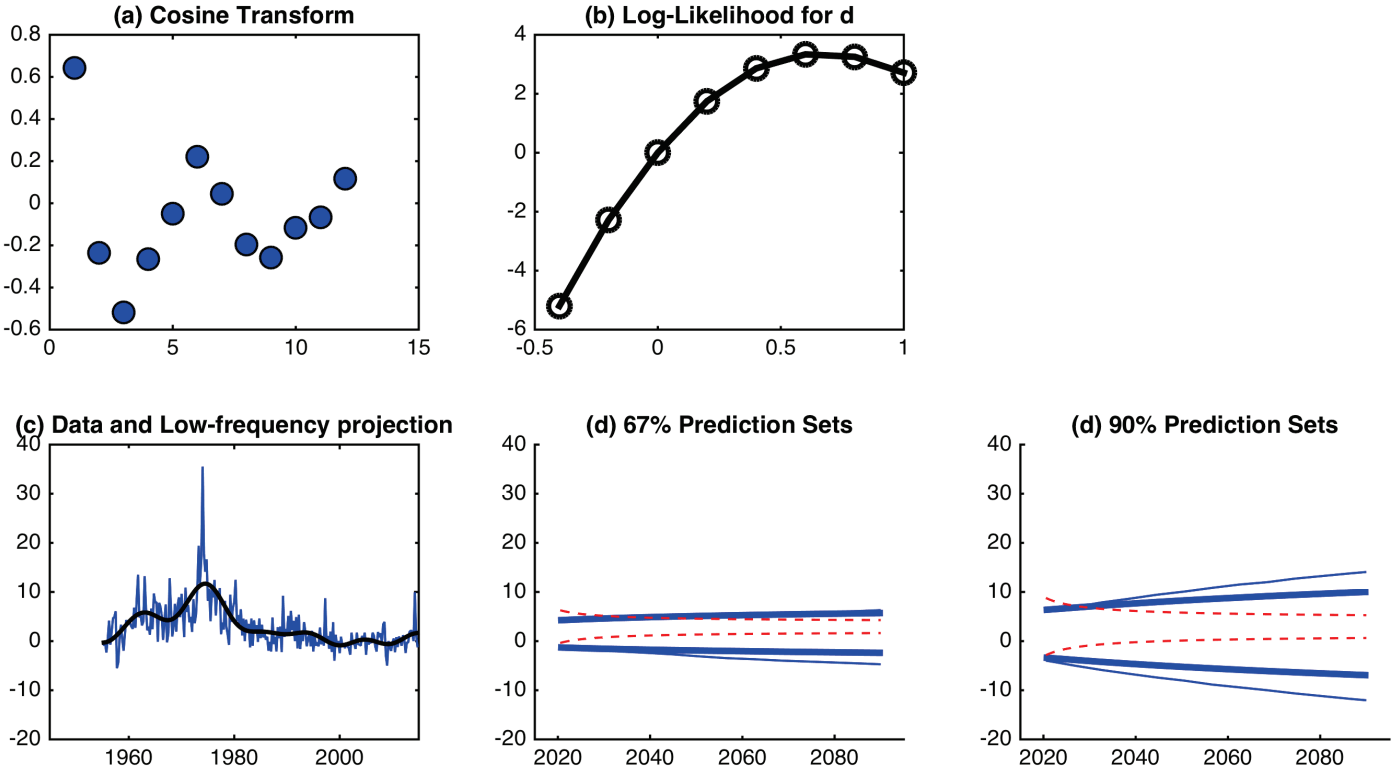
Sample period: 1947:Q2 – 2014:Q4, $\bar{x}_{1:T} = 3.53$, $s_{LR} = 10.38$



See notes to C.1.

C.8 Inflation (CPI, Japan)

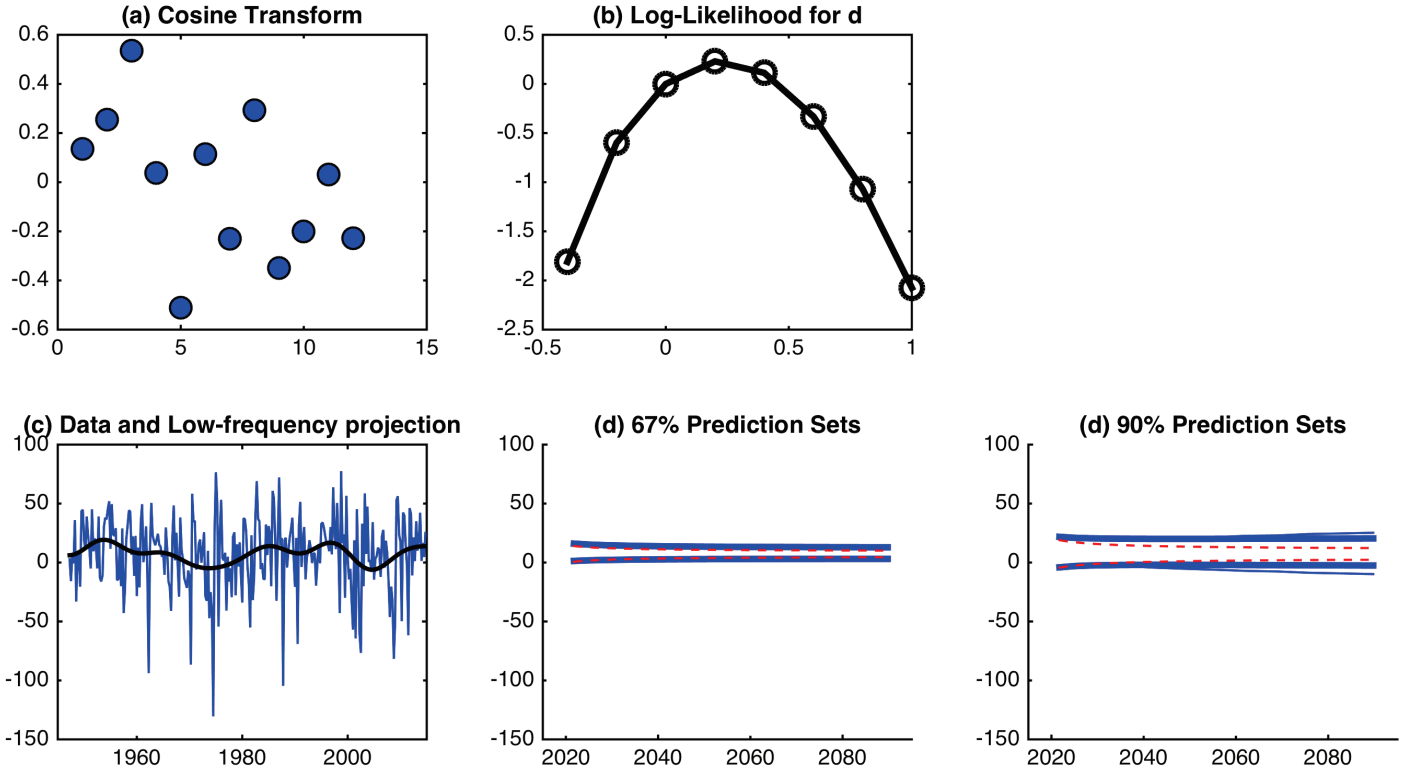
Sample period: 1955:Q2 – 2014:Q4, $\bar{x}_{1:T} = 2.96$, $s_{LR} = 15.02$



See notes to C.1.

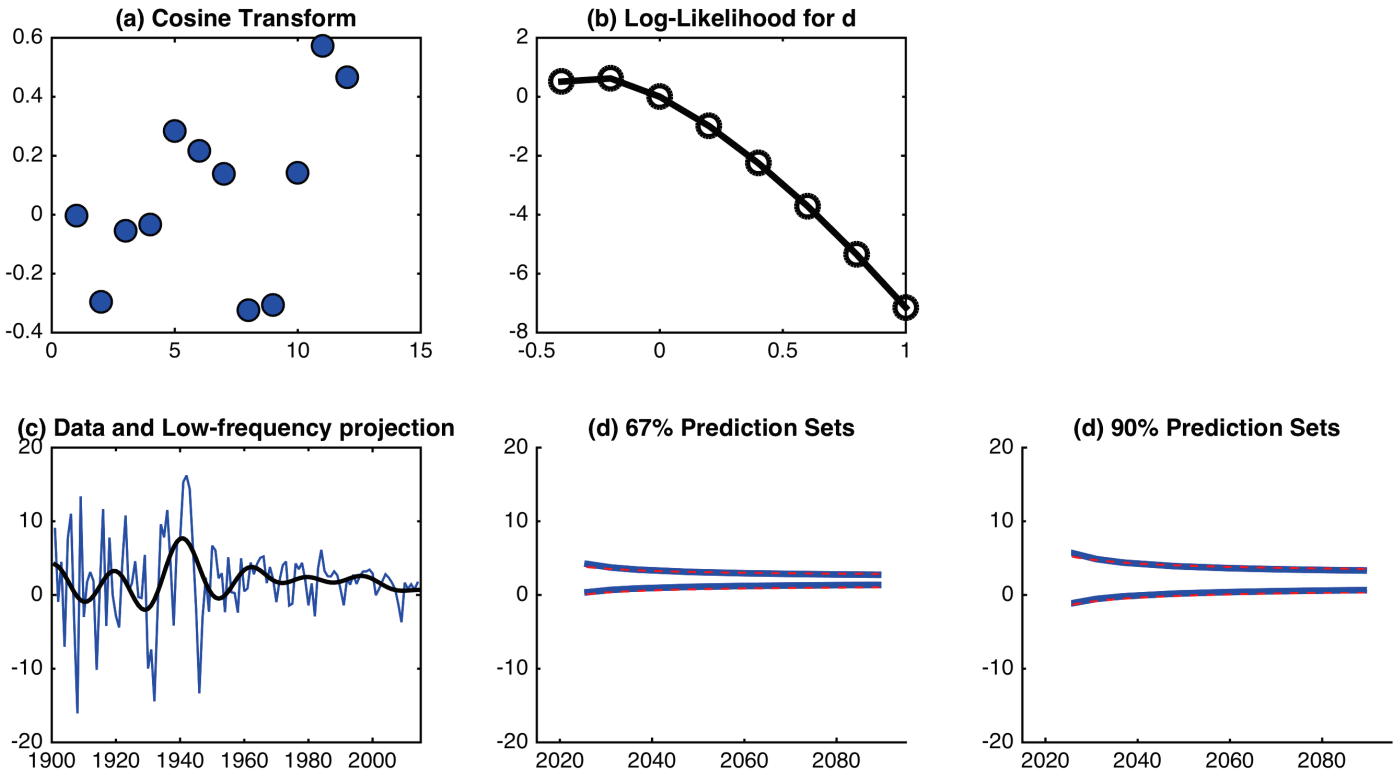
C.9 Stock Returns

Sample period: 1947:Q2 – 2014:Q4, $\bar{x}_{1:T} = 7.30$, $s_{LR} = 33.11$



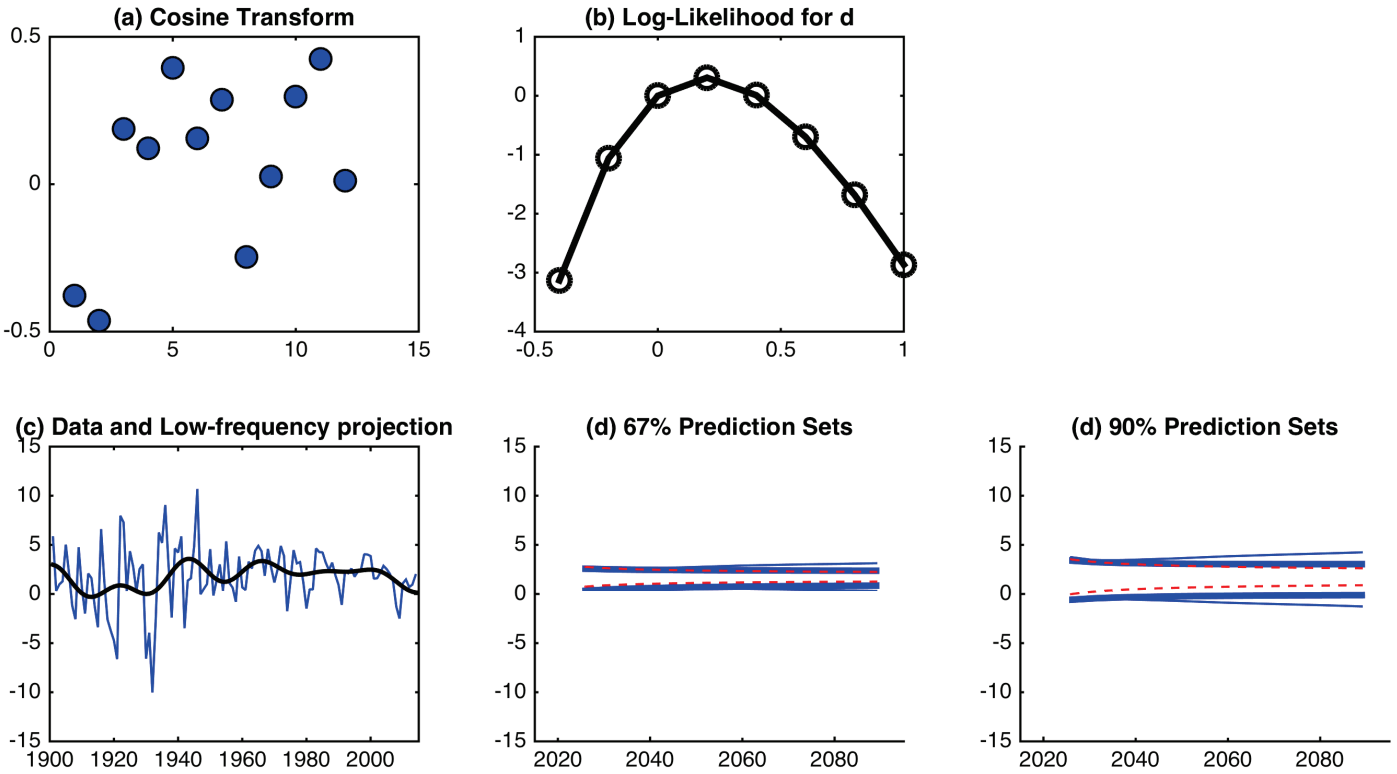
See notes to C.1.

C.10 Growth rate of real per-capita GDP (annual data)
Sample period: 1901 – 2014, $\bar{x}_{1T} = 1.95$, $s_{LR} = 5.99$



See notes to C.1.

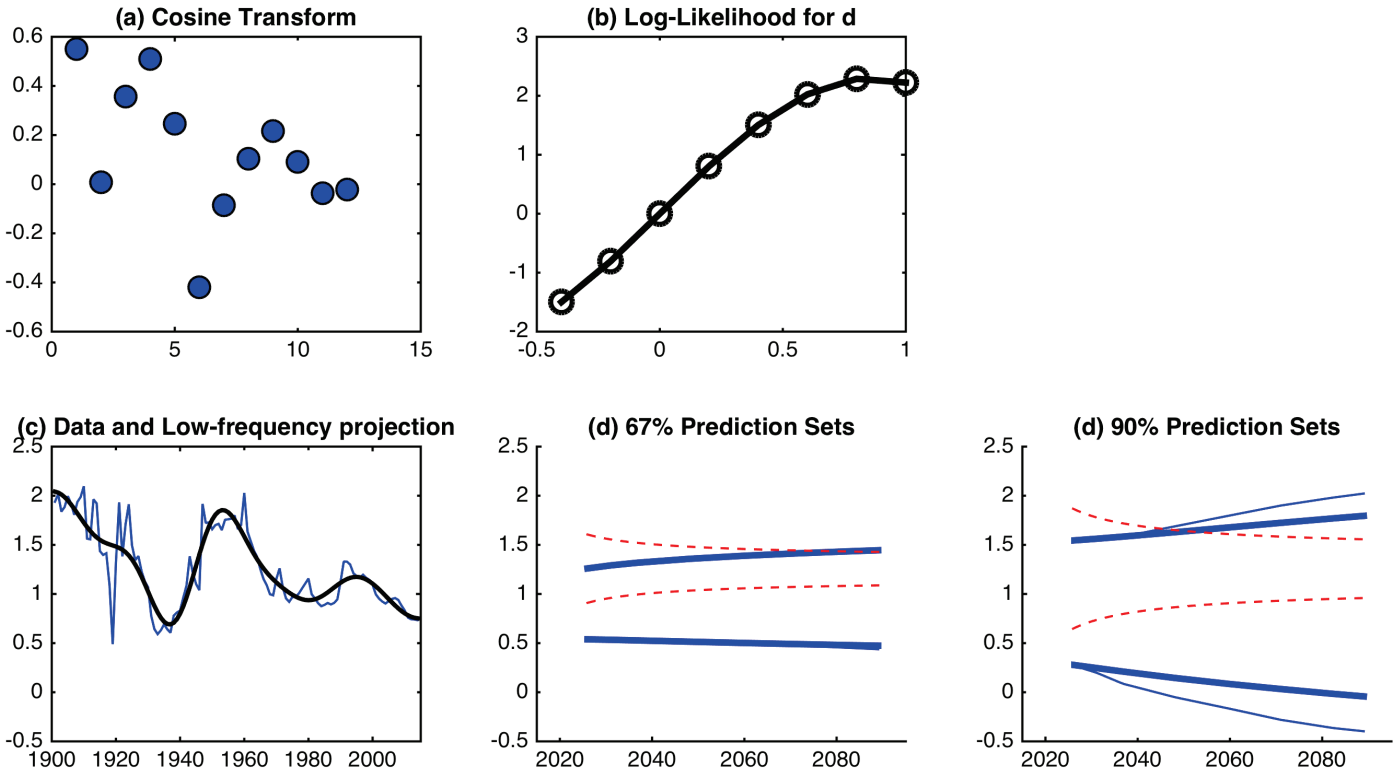
C.11 Growth rate of real per-capita consumption expenditures (annual data)
Sample period: 1901 – 2014, $\bar{x}_{1:T} = 1{,}76$, $s_{LR} = 3.27$



See notes to C.1.

C.12 Growth rate of population (annual data)

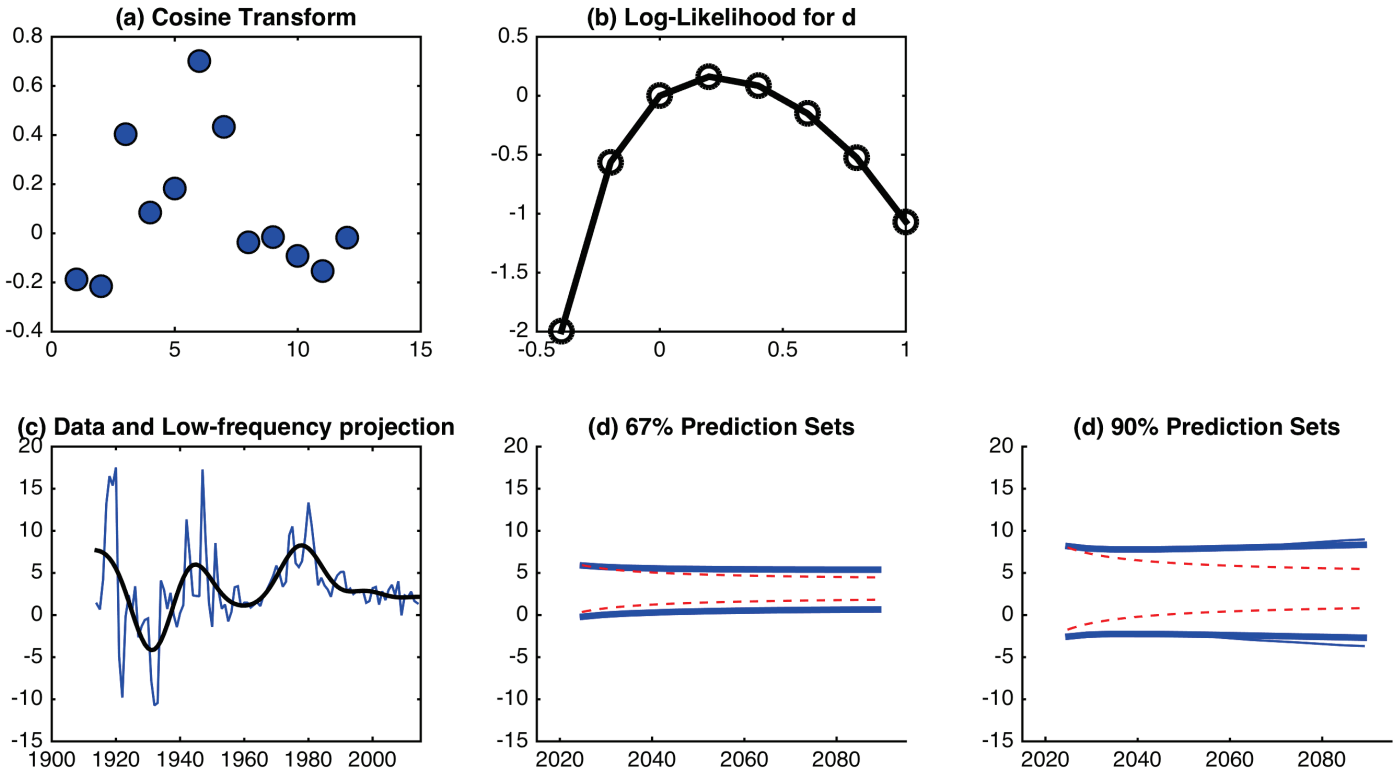
Sample period: 1901 - 2014, $\bar{x}_{1:T} = 1.26$, $s_{LR} = 1.13$



See notes to C.1.

C.13 Inflation (CPI, annual data)

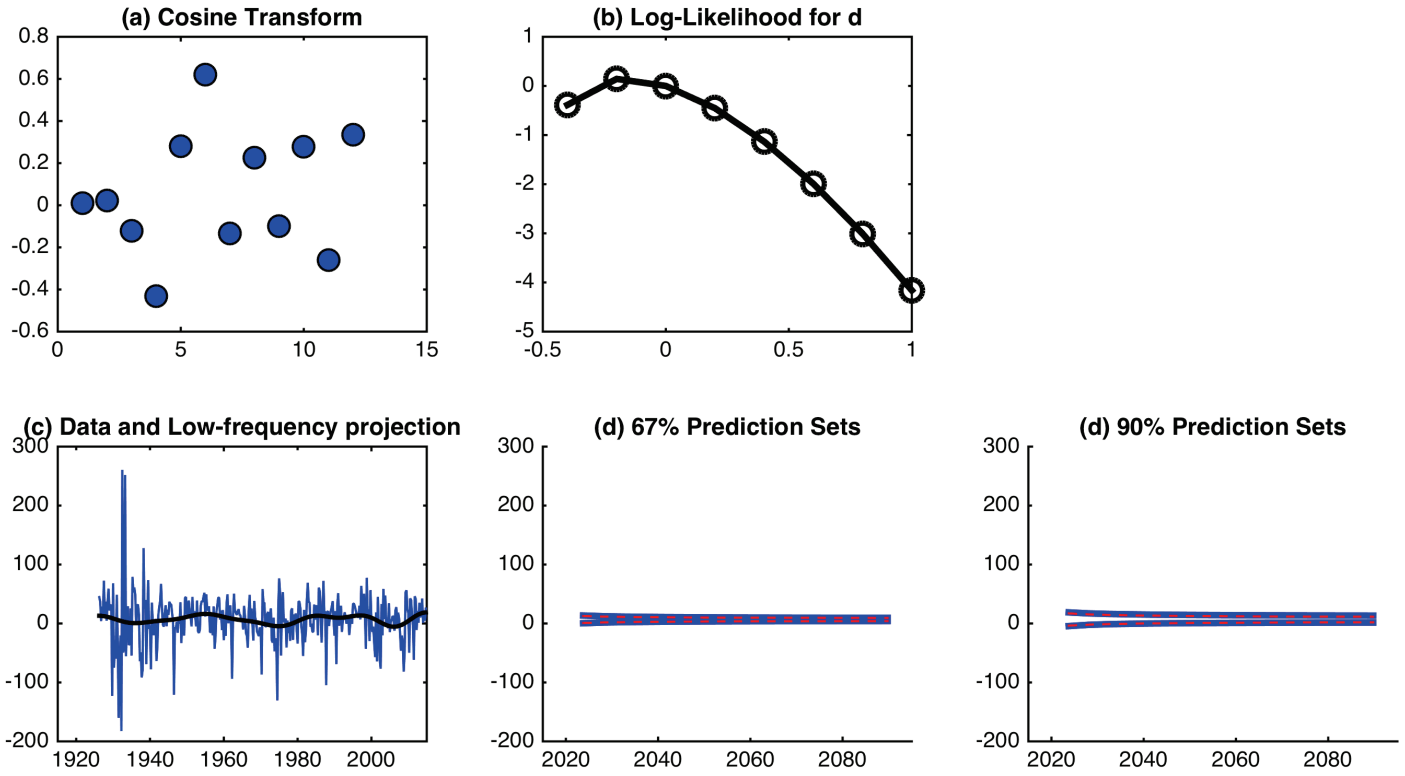
Sample period: 1914 - 2014, $\bar{x}_{1:T} = 3.15$, $s_{LR} = 8.56$



See notes to C.1.

C.14 Stock Returns

Sample period: 1926:Q2 – 2014:Q4, $\bar{x}_{1:T} = 6.72$, $s_{LR} = 33.56$



See notes to C.1.