Annals of Actuarial Science, Vol. 11, part 1, pp. 100–127.
doi:10.1017/S174849951600021X
First published online 21 November 2016

Yet more on a stochastic economic model: Part 3B: stochastic bridging for retail prices and wages

A. D. Wilkie*
InQA Limited, Dennington, Ridgeway, Horsell, Woking GU21 4QR, UK
Şule Şahin
Department of Actuarial Sciences, Hacettepe University, 06800 Ankara, Turkey

Abstract
This is the second subpart of three in a long paper in which we consider stochastic interpolation for the Wilkie asset model, considering both Brownian bridges and Ornstein–Uhlenbeck (OU) bridges. In Part 3A, we developed certain properties for both these types of stochastic bridge, and we investigate the properties of many of our data series on the same lines. We have several economic or investment series, which all have their own peculiarities. In this paper, we cover only retail prices and wages. The other series are dealt with in Part 3C. We find that, although the annual series for the rate of inflation is generated by an AR(1) model, which is the discrete time equivalent of an OU process, an OU bridge is not suitable. We need to use a Brownian bridge on the logarithm of the Price Index. Further, the standard deviation of the monthly increments in any year is, as we find empirically from the data, a function of the change in the annual value, and further there is correlation between the monthly increments in successive years.

Keywords
Wilkie model; Stochastic interpolation; Bridging; Retail prices; Wages

1. Introduction

1.1. In Parts 1 and 2 of this series of papers (Wilkie *et al.*, 2011; Wilkie & Șahin, 2016), we updated the Wilkie model (see Wilkie 1986 and Wilkie 1995) to 2009 and described several aspects of using such an annual model. In the three subparts of this long paper, we consider ways to make the model applicable to shorter time steps, by stochastic interpolation between the points of a stochastically generated annual model.

1.2. In Part 3A, we described the basic ways of doing stochastic interpolation by the familiar Brownian bridges, and rather less familiar Ornstein–Uhlenbeck (OU) bridges and went into the properties of these forms of bridging. In this paper, Part 3B, we consider models for retail prices and for wages within the Wilkie model and how stochastic interpolation can be applied to these series. We find that what seem to be the obvious methods are not always satisfactory. We also investigate the source data for each series, in so far as monthly data are available, and we find that the actual monthly data generally does not behave in the same way as the annual data. We produce plausible models in each case for monthly modelling.

*Correspondence to: A. D. Wilkie, InQA Limited, Dennington, Ridgeway, Horsell, Woking GU21 4QR, UK. Tel. +441483 725984 or 01483 725984; E-mail: david.wilkie@inqa.com

1.3. We investigate stochastic bridging for retail prices in sections 2, 3 and 4, and for wages in sections 5, 6 and 7. We find that the obvious method which conforms with the underlying annual models, that of applying an OU bridge to the annual rate of inflation, $I(t)$, or the annual growth of wages, $J(t)$, is not satisfactory, and it is only after several alternative possibilities have been tried and found wanting that we can settle on a final satisfactory method. This turns out in each case to be a Brownian bridge with a variable monthly standard deviation and with correlation between monthly innovations over successive years for each series.

1.4. In order to check the behaviour of any model, it is essential to simulate projections of the future experience in both a deterministic way (with zero innovations, giving the expected values) and stochastically. The annual model can be projected in either way, and the monthly model within each of these in either way, giving us four possibilities. All should be investigated, as we show below.

2. The Inflation Model

2.1. In the first place we consider the model for the Retail Prices Index (RPI), $Q(t)$, and the rate of inflation, $I(t) = \ln Q(t) - \ln Q(t-1)$. The values of $I(t)$ are included in the models for almost every other series, so it is fundamental in the simulations and therefore it is desirable to find a satisfactory method for $Q(t)$ and $I(t)$ first. But, as we shall show, there are considerable problems in stochastic interpolation for this variable.

2.2. The basic formulae for $I(t)$ and $Q(t)$ are

$$QE(t) = QSD.QZ(t)$$

$$IN(t) = QA.IN(t-1) + QE(t)$$

$$I(t) = QMU + IN(t)$$

$$QL(t) = QL(t-1) + I(t)$$

$$Q(t) = \exp(QL(t))$$

Parameters suggested in Part 1 (Wilkie *et al.*, 2011) were

$$QMU = 0.043;\ \ QA = 0.58;\ \ QSD = 0.04.$$

2.3. The model for $I(t)$ is a simple AR(1) model, so the obvious first step is to use OU bridges for $I(t)$. We then wish to calculate interpolated values of $Q(t)$, or rather of $QL(t) = \ln Q(t)$. But to calculate a value for $Q(0, m)$ in month m of year 0, we need $Q(-1, m)$, because we would put $QL(0, m) = QL(-1, m) + I(0, m)$. We may have actual observed monthly values of Q for the past year, but we might not have them. If we do not have them, a neutral sort of start would be to use a Brownian bridge, or just linear interpolation, between the observed (or assumed) $Q(-1)$ and $Q(0)$, and we do this for the time being. If we use the actual values, we need to be careful, because we do not wish to replicate the seasonal pattern of $I(t)$ in the year $(-1, 0)$ in the simulations for future years.

2.4. In fact we find that the obvious method, which we denote Method 1, using OU bridges for $I(t)$, does not give a satisfactory result. We can show this most clearly by using very extreme starting points for $I(0)$. Figure 1 shows a 10-year deterministic simulation for $I(t)$ using the parameters noted above, and $I(0) = 0.40$. We get the usual exponential declining asymptotically towards $I(\infty) = 0.04$, with the same curve fitting both the annual and the monthly values.

https://doi.org/10.1017/S174849951600021X Published online by Cambridge University Press

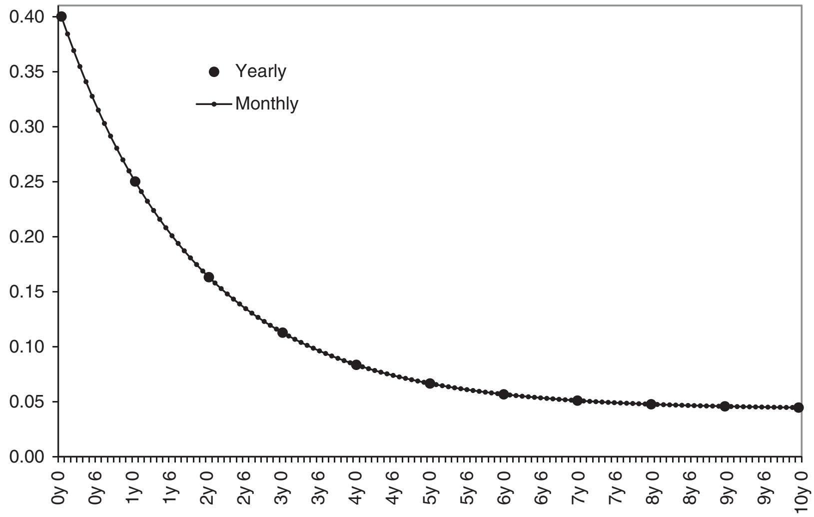

Figure 1. Deterministic simulation for 10 years of $I(t)$, Method 1, with $I(0) = 0.40$.

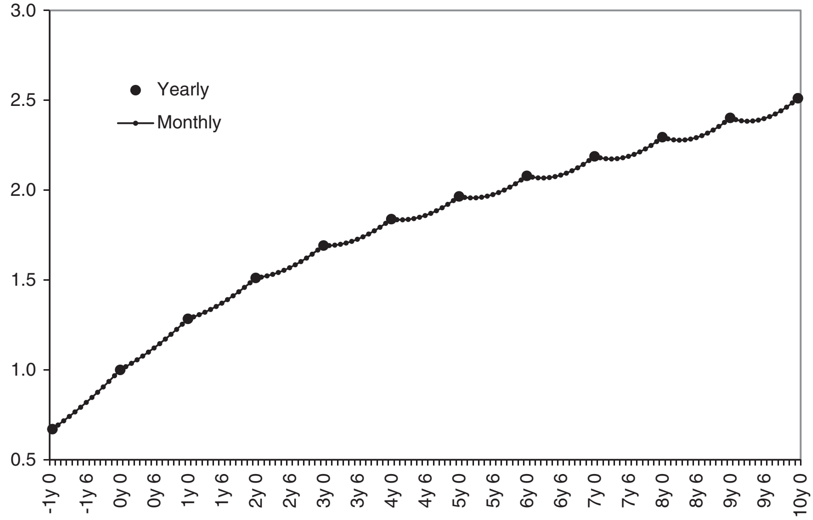

Figure 2. Deterministic simulation for 10 years of $Q(t)$, Method 1, with $I(0) = 0.40$.

2.5. In Figure 2 we show the corresponding values of $Q(t)$, also deterministic and also for 10 years, but including the year from −1 to 0. The deterministic simulation for a Brownian bridge is simply linear interpolation, but this has been carried out on $QL(t)$ and here we show $Q(t)$, so for the year from −1 to 0, the graph of $I(t)$ has a very gentle exponential curve rather than a straight line. But as the years increase we get "scallops" forming, increasingly large, with intermediate values of $Q(t, j)$ falling below that for $Q(t)$. In Figure 3, we continue up to 50 years, and we see the scallops getting bigger. It can be shown algebraically that the scallops are the consequence of the curved sections of the graph of $I(t)$, which are seen in Figure 1, and that they accumulate in the graph of $Q(t)$, reaching an asymptotic upper limit.

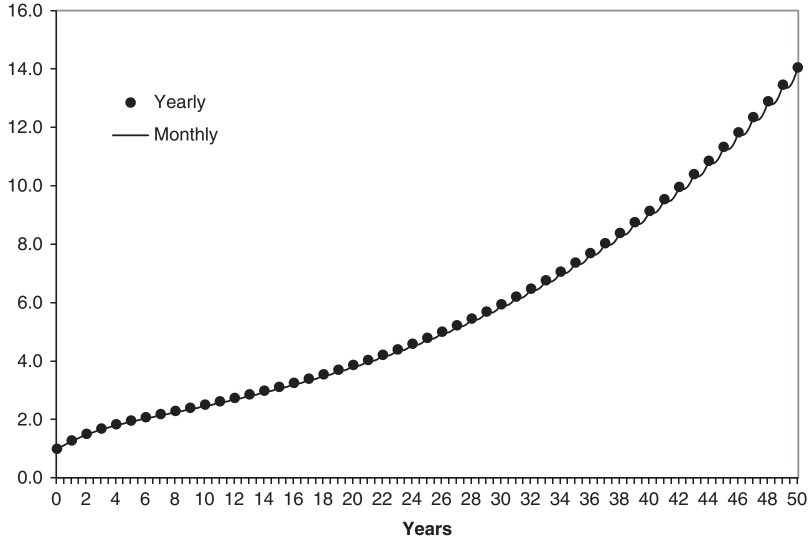

Figure 3. Deterministic simulation for 50 years of $Q(t)$, Method 1, with $I(0) = 0.40$.

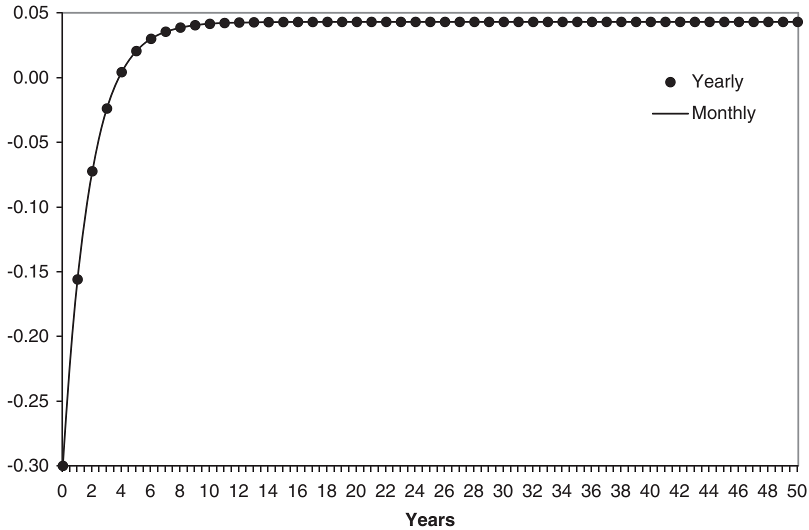

Figure 4. Deterministic simulation for 50 years of $I(t)$, Method 1, with $I(0) = -0.30$.

2.6. If we start with an extremely low value of $I(0)$, for which we use -0.3, we get the results shown in Figures 4 for $I(t)$ and 5 for $Q(t)$, both for 50 years. The value of $I(t)$ tends reasonably quickly towards its asymptotic value of $QMU = 0.043$, but the graph of $Q(t)$ shows scallops the other way up, with arches getting steadily larger. It is clear that the obvious first method (Method 1) is unsatisfactory.

2.7. We try, therefore, Method 2, using Brownian bridges for $I(t)$. This means that the deterministic simulation for $I(t)$ shows the annual points joined by straight lines, and if we start with a Brownian

https://doi.org/10.1017/S174849951600021X Published online by Cambridge University Press

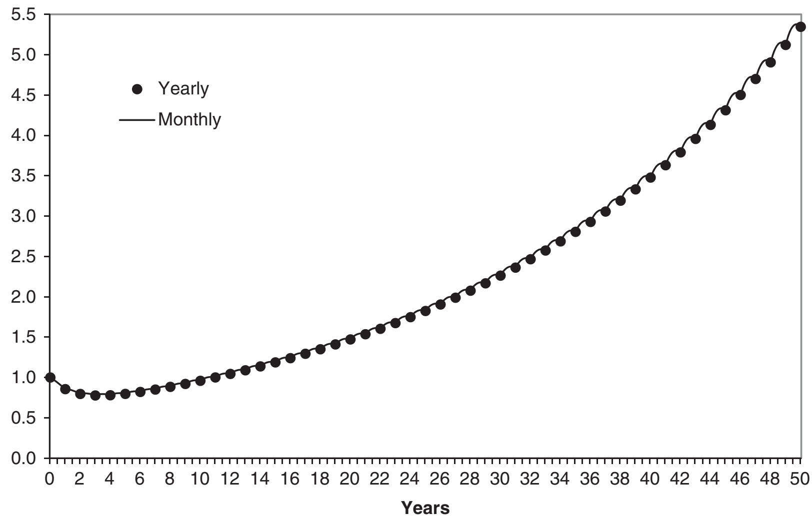

Figure 5. Deterministic simulation for 50 years of $Q(t)$, Method 1, with $I(0) = -0.30$.

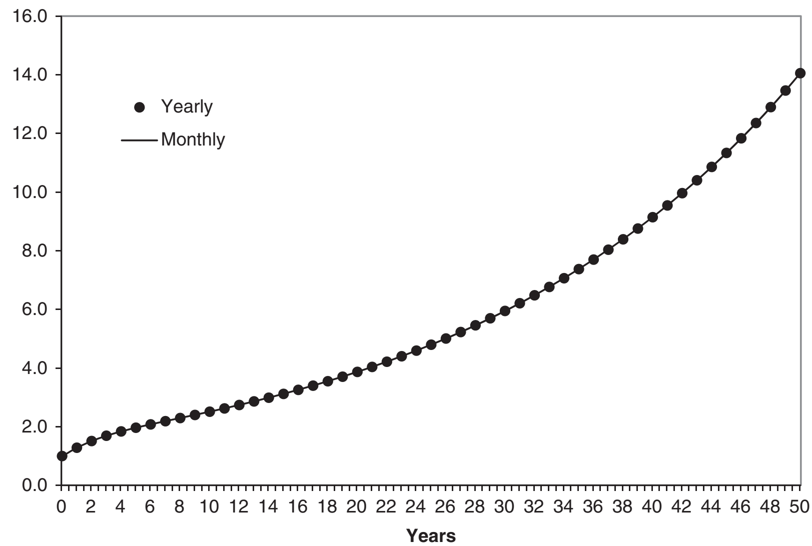

Figure 6. Deterministic simulation for 50 years of $Q(t)$, Method 2, with $I(0) = 0.40$.

bridge for $\ln Q(t)$ from –1 to 0, then the deterministic monthly simulations for $\ln Q(t)$ are also straight lines, although for $Q(t)$ these appear as slightly exponentially curved, as we show in Figure 6. This looks more plausible than the scallops in Figures 3 and 5.

2.8. We now try a stochastic simulation for the monthly values, keeping the annual ones deterministic, applying the Brownian bridges to I and calculating the values of Q from the simulated values of I. One set of simulated values for $I(t)$ is shown in Figure 7, and the corresponding values of $Q(t)$

https://doi.org/10.1017/S174849951600021X Published online by Cambridge University Press

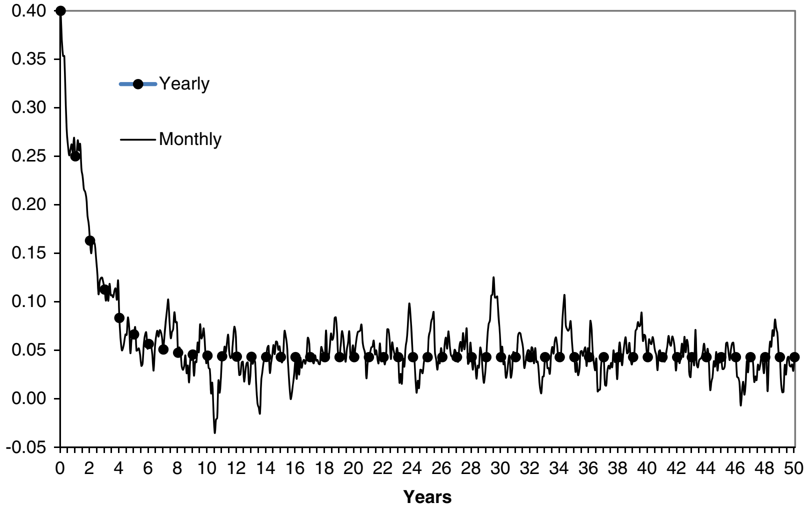

Figure 7. Stochastic monthly with deterministic yearly simulation for 50 years of $I(t)$, Method 2, with $I(0) = 0.40$.

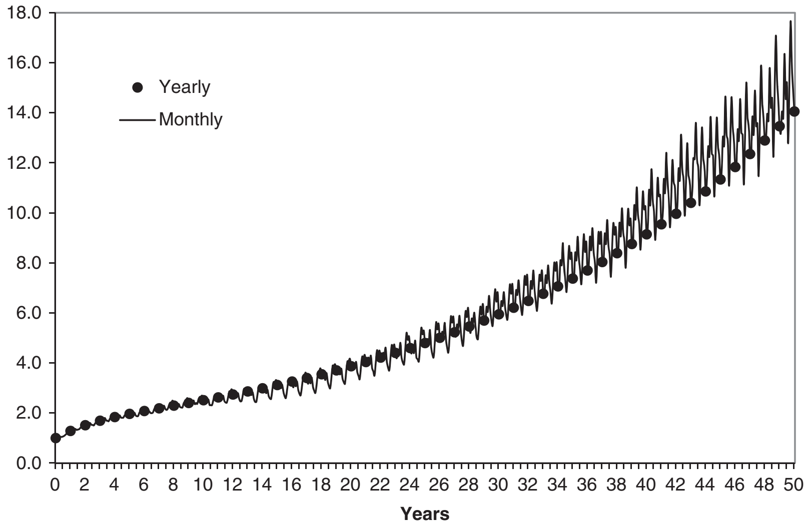

Figure 8. Stochastic monthly with deterministic yearly simulation for 50 years of $Q(t)$, Method 2, with $I(0) = 0.40$.

are shown in Figure 8. The values of $\ln Q(t)$ for the year from −1 to 0 have been simulated by a Brownian bridge, and this causes no problems. The simulated values of $I(t)$ are plausible – but we would still have to decide on a suitable value for σ_m, the monthly standard deviation; we have used $0.011547 = 0.04/\sqrt{12}$, provisionally.

2.9. An obvious problem arises, however, as time increases. The simulated values of $Q(t)$ oscillate more and more with increasing t, and show a similar, but not uniform, pattern from year

https://doi.org/10.1017/S174849951600021X Published online by Cambridge University Press

to year. This is because $QL(t, j)$ is calculated from the sum of the values of $I(s, m)$ for years $s = 0$ to t:

$$QL(t,j) = QL(-1,j) + \sum_{s=0,t} I(s,j)$$

This means than any value of $I(s, m)$ continues its effect into all subsequent values of $QL(t, m)$. Further, the variance of $QL(t, m)$ depends on the sum of the variances of the preceding values of $I(s, m)$, so that it increases with t. Method 2 is not a satisfactory method either, because of the increasing variance of $QL(t, m)$.

2.10. We try, therefore, Method 3. We simulate monthly values of $\ln Q(t)$ using Brownian bridges, and calculate the monthly values of $I(t)$ from these. But what value should we use for the standard deviation, σ_m? If $QL(t)$ is calculated using a Brownian bridge with some value for σ_m, so that the forwards deviations defined in Part 3A are based on this value, then the values of changes in $I(t)$ have twice the variance of changes in $QL(t)$. We have

$$I(t,j) = QL(t,j) - QL(t-1,j)$$

and

$$I(t,j+1) = QL(t,j+1) - QL(t-1,j+1)$$

$$\text{So } I(t,j+1) - I(t,j) = \{QL(t,j+1) - QL(t,j)\} - \{QL(t-1,j+1) - QL(t-1,j)\}$$

If the variance of the monthly change in $QL(t, j)$ is σ_m^2, then the variance of the monthly change in $I(t, j)$ is $2\sigma_m^2$.

2.11. At the start of any year t, the "forecast value" of $I(t+1)$ has standard deviation of QSD. Then the forecast value of $QL(t+1)$, which equals the known $Q(t)$ plus the forecast $I(t+1)$ also has standard deviation QSD. It depends on what values we assume are known, and which we condition on. Another approach is to revert to the data, for which we do have monthly values, and we discuss this next.

3. Analysing the Inflation Data Series

3.1. We have available published monthly values of the Retail Prices Index, or its predecessor, from August 1914. We show in Figure 9 values of the RPI, $Q(t)$ (multiplied by 2.5 to fit the scale better) from December 1922 to December 2014, which includes the period which we consider for analysis, and in Figure 10 values of the annual change in $\ln Q(t)$, $I(t)$, for the same period. Our simulations need to resemble these graphs adequately, and we use this as an important and necessary criterion throughout. Using a visual comparison is not a proof that the series have the right mathematical properties, but failure to look right, as for example in Figures 2 and 8, may be clear evidence of an unsatisfactory model.

3.2. The RPI series that we use began in August 1914 with a value of 100, but the published values in the first few years were multiples of 2.5. So in any month, the change was restricted to 0, ±2.5 or multiples thereof. So monthly changes fall into a very lumpy discrete series. From April 1920, when the value had reached 232, the values were rounded to single integers, but even so the available monthly changes were of the order of multiples of ½%. Because we take logarithms the monthly difference are not on a linear grid, but they are not very like a continuous distribution. In our earlier

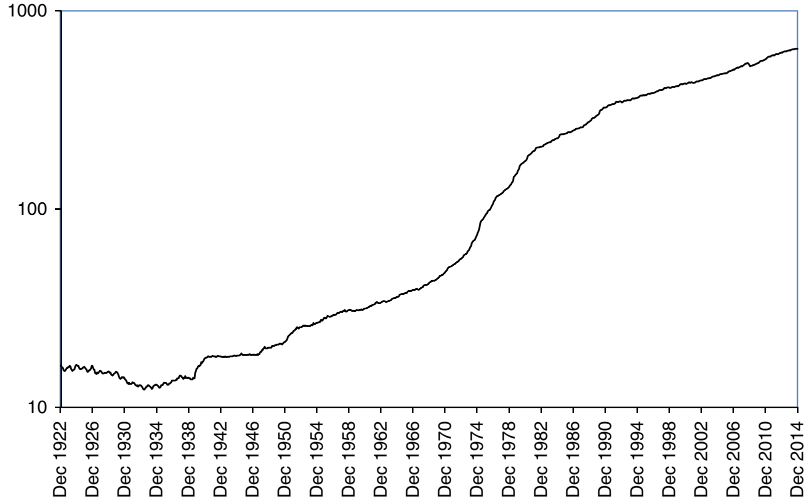

Figure 9. Values of $2.5 \times Q(t)$, the RPI, monthly, from December 1922 to December 2014.

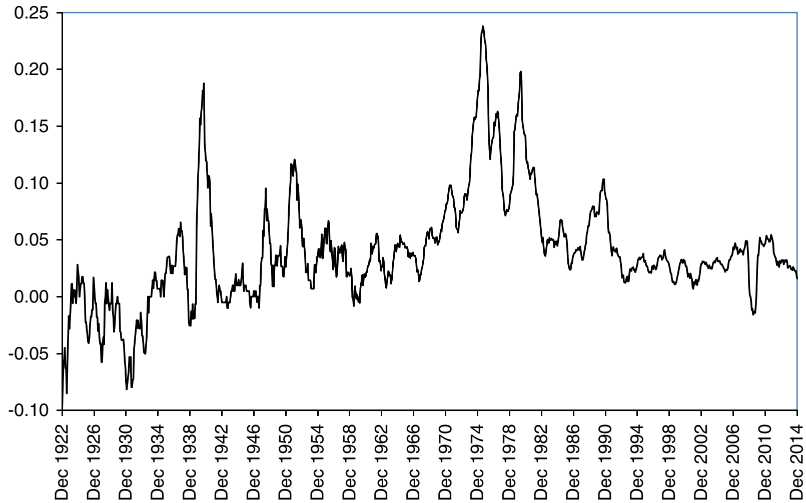

Figure 10. Values of $I(t)$, monthly, from December 1922 to December 2014.

analysis, we have used the change over one year to give values of $I(t)$, and this lumpiness is not so severe. In June 1947, the index, now 203, was replaced by a new one with a value of 100, but that too was rounded to integers, and it was not until February 1956 that a third series started, also at 100, but with one decimal point given. From then on, the differences of the logarithms are more smoothly distributed. The index was rebased at 100 in January 1962, again in January 1974, and again in January 1987, but in each case one decimal place has been retained.

3.3. We demonstrate the effect of this rounding by calculating the monthly changes of $\ln Q(t)$, and we plot them in Figure 11, from 1923 to 2014. The lumpiness in the early years is obvious, with the dots

https://doi.org/10.1017/S174849951600021X Published online by Cambridge University Press

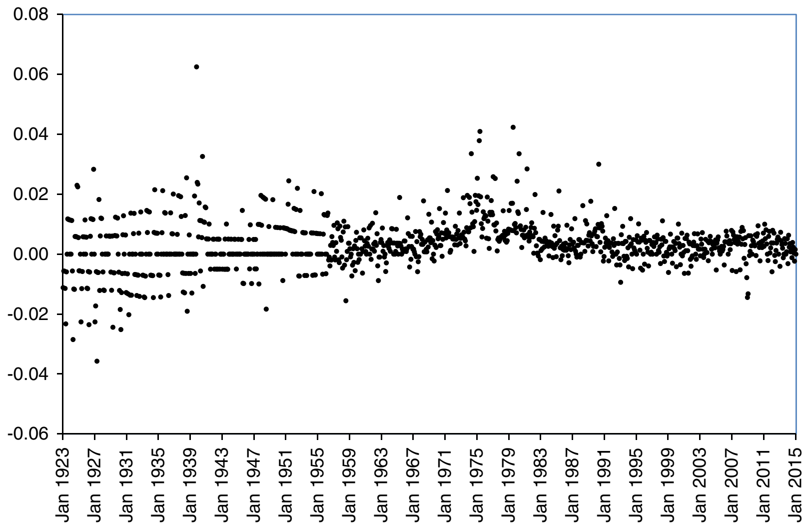

Figure 11. Monthly changes in ln$Q(t)$ from 1923 to 2014.

lying on fluctuating lines. The change in 1947 is also visible, because the index was rebased from 203 to 100, so the available monthly changes became proportionately bigger. After 1956, one decimal place is given in the index and the series looks more reasonable.

3.4. In our analysis of the actual data, we have generally calculated figures for the whole period, from June 1923 to June 2014, and for two sub-periods: (1) from June 1923 to June 1956 and (2) from June 1956 to June 2014. We have also taken the data for the later subperiod, which includes all the series for which one decimal place is given, and we have rounded the source values to integers, and constructed a new chain-linked index, where the linking factors are slightly different from those for the original index. We call the original series (2A), the new one (2B). We can then look at the differences between them as well as at the differences between periods (1) and (2).

3.5. For each of the periods, we can analyse the logarithms, $QL(t, j)$ for year t month j, which form the fundamental series, or we can analyse the annual inflation rates, $I(t, j)$, or both. But the monthly change in the annual rate from $I(t, j)$ to $I(t, j+1)$ consists of the change in QL in month $j+1$ of year t minus the change in QL in the same month in the previous year. It is satisfactory to use non-overlapping annual rates of inflation, as we do for the main model, but there may be problems in analysing these overlapping annual rates. So we concentrate first on analysing QL.

3.6. We have used the period from June 1923 onwards to fit the main model, so it is consistent to use the period from June 1923 for investigating the bridging. But as at the time of writing we have data up to June 2014, so we use that, giving a total of 91 years of data. We assume that the June values are for month 0 in any year, and are fixed, and we count months through to the following June. We do not wish here to analyse all features of the underlying series, which we assume has given us our annual model; we wish only to consider the bridging process over each year between the June values. We also analyse the two sub-periods (1) and (2), the latter in versions (2A) and (2B).

https://doi.org/10.1017/S174849951600021X Published online by Cambridge University Press

3.7. We analyse these QL values as if the path from one June to the next was a Brownian bridge. Given the fixed annual values corresponding to the $X(t, 0)$ values of Part 3A, we can calculate the expected values over the year, the "sideways deviations" over the months of the years, and the "forwards deviations", all as described in Part 3A. We can calculate the means and standard deviations of these Ds_j and Df_j values for all years for each month separately. We can then test whether these monthly means are near enough to zero, and whether the standard deviations have the shape and size that would allow us to assume that they behave as if they had been produced by Brownian bridges, or not. We calculate the variance and standard deviation within each year separately, from the twelve forwards deviations in the year. Their sum is necessarily zero, and our estimate of the standard deviation for the year, $\sigma_m(t)$, is obtained by dividing the sum of squares by 11, as described in section 8 of Part 3A. We get an estimate of the overall variance and standard deviation by summing all the squares and dividing by 91×11. This gives a different answer from the average of the annual standard deviations, $\sigma_m(t)$, but it agrees with the average of the annual variances. We standardise the forwards deviations in each year by dividing each by $\sigma_m(t).11/12$. We then look also at the inter-year correlations of the standardised deviations, and later at cross-correlations with other series in corresponding months.

3.8. In Table 1, we show statistics for the forwards deviations of QL over the 91 years; we have multiplied the means and standard deviations by 100 to scale them more conveniently. We calculate the standard errors of the monthly means by dividing the monthly standard errors by $\sqrt{91}$, and from that we calculate T-ratios for each month. Only six of these T-ratios are not significant, assuming significance to be outside the range (−2, +2). The deviations (i.e. the increase in each month minus the annual mean change) were significantly high in October, November and April, and significantly low in August, November and especially January, for which the T-ratio is −5.13. There is clearly a monthly variation in the changes – though not a uniform seasonal one. We note that the sum of the Df values in each year is zero, so the sum of the means is also zero.

3.9. In the same table, we show the skewness and kurtosis of the monthly values. Both show much higher values than they would if they were normally distributed. Only the values for September, 0.12 and 3.96, are plausibly normal compared with the expected values of 0 and 3.0, but Jarque–Bera tests on all the others show highly significant non-normality, many with tiny p-values.

Table 1. Statistics for forwards deviations, 100 Df_j, of QL, from June 1923 to June 2014.

Months	$100 \times$ mean of Df_j	$100 \times$ s.d. of Df_j	S.E. of Mean	T-ratio	Skewness	Kurtosis
July	0.0176	0.9084	0.0952	0.19	1.11	3.82
August	−0.1982	0.6964	0.0730	−2.71	−0.45	3.88
September	−0.0138	0.4894	0.0513	−0.27	0.12	3.96
October	0.2306	0.7335	0.0769	3.00	3.19	18.71
November	0.2121	0.6144	0.0644	3.29	1.72	7.55
December	0.0197	0.4346	0.0456	0.43	0.38	4.67
January	−0.2665	0.4956	0.0520	−5.13	0.05	4.23
February	−0.0447	0.4760	0.0499	−0.89	−0.61	4.14
March	−0.1162	0.4390	0.0460	−2.52	−0.74	4.09
April	0.2567	1.0970	0.1150	2.23	−1.15	4.19
May	−0.0473	0.5513	0.0578	−0.82	0.95	6.95
June	−0.0502	0.4694	0.0492	−1.02	0.79	4.32
Mean	0.0					

https://doi.org/10.1017/S174849951600021X Published online by Cambridge University Press

Table 2. Statistics for forwards deviations, 100 Df_j, of QL, from June 1923 to June 1956, and from June 1956 to June 2014.

	1923–1956				1956–2014			
Months	Mean of Df_j	T-ratio	Skewness	Kurtosis	Mean of Df_j	T-ratio	Skewness	Kurtosis
July	0.7593	4.48	−0.34	3.10	−0.4043	−5.98	3.40	21.95
August	−0.3496	−2.07	0.14	2.43	−0.1120	−1.88	−1.25	5.06
September	0.0959	0.82	−0.25	2.80	−0.0762	−1.76	0.01	2.56
October	0.5897	3.22	2.00	8.67	0.0262	0.64	0.56	3.52
November	0.6586	4.96	0.83	4.33	−0.0420	−1.11	0.00	3.14
December	0.1567	1.73	0.55	3.04	−0.0582	−1.25	−0.70	4.96
January	−0.4046	−4.77	−0.65	4.93	−0.1879	−2.96	0.49	3.26
February	−0.2658	−2.50	0.11	3.06	0.0812	1.96	−0.29	3.43
March	−0.3558	−3.39	0.32	2.56	0.0202	0.73	−0.21	3.56
April	−0.6997	−3.43	−0.39	2.18	0.8009	11.31	0.48	3.59
May	−0.1469	−1.13	0.65	3.47	0.0093	0.18	2.82	17.41
June	−0.0378	−0.33	0.59	2.68	−0.0573	−1.41	0.82	3.53

This may have upset the T-ratio test we have used above, but even so the variation in monthly means is large.

3.10. Inspection of the values of Df in individual months shows that there were many quite large values, in both directions, in the early years. However, in about the last 50 years, from 1964 onwards, the large positive values have often occurred in April, presumably because of an increase in taxes announced in the government's budget at that time.

3.11. We show some of the corresponding figures for the two sections, 1923–1956 and 1956–2014, in Table 2. We can see that the monthly averages (again multiplied by 100) give rather different results. In the first period, seven months have T-ratios absolutely greater than 3.0; in the second period only two have; July is very low, April very high, but the reverse applies to those months in the earlier period. The standard deviations (not shown) are all much larger for the first period, and also the standard errors of the mean are based on fewer years (33 instead of 58).

3.12. We have the same statistics for period (2B), and can compare them with those in Table 2 for 1956–2014 (2A). The standard deviations for (2B) are all larger than for (2A), but by not nearly as much as those for (1) exceed those for (2A). Consequently the T-ratios and the skewness and kurtosis are a bit smaller. The means are sometimes larger, sometimes smaller.

3.13. The estimated monthly standard deviation σ_m over the whole period is 0.006969, equivalent to an annual standard deviation σ_y of 0.0241. For the two separate periods, the estimates of σ_m are 0.009548 and 0.004933, equivalent to annual standard deviations σ_y of 0.0331 and 0.0171. For (2B) they are: σ_m 0.005613 and σ_y 0.0194, larger than for (2A). The figures for (1) and (2A) are very different, but all the estimates of σ_y are much lower than the value of QSD in the annual model, estimated in 2009 at close to 0.04.

3.14. We can observe that, in general, the skewness and kurtosis of the monthly values for the two sections are less than for the period as a whole, except for some exceptionally large values for July

Table 3. Statistics for sideways deviations, 100 Ds_j, of QL, from June 1923 to June 2014.

Months	Mean of Ds_j	s.d. of Ds_j	T-ratio	Multiple	s.d. ratio, rs_j	M-ratio
June	0	0	0	0	0	0
July	0.0176	0.9084	0.19	1.3035	0.9574	1.36
August	−0.1805	1.0295	−1.67	1.4773	1.2910	1.14
September	−0.1943	1.2920	−1.43	1.8539	1.5000	1.24
October	0.0363	1.4912	0.23	2.1398	1.6330	1.31
November	0.2483	1.9019	1.25	2.7292	1.7078	1.60
December	0.2681	2.0476	1.25	2.9381	1.7321	1.70
January	0.0016	1.8418	0.01	2.6428	1.7078	1.55
February	−0.0430	1.6387	−0.25	2.3513	1.6330	1.44
March	−0.1592	1.4443	−1.05	2.0725	1.5000	1.38
April	0.0975	0.6682	1.39	0.9588	1.2910	0.74
May	0.0502	0.4694	1.02	0.6736	0.9574	0.70
June	0	0	0	0	0	0

and May in the second period, and for October in the first. These may be attributable to some very large values, in recent years perhaps because of large changes in the rate of value added tax.

3.15. We show, for the whole period, the means of the values of the sideways deviation, Ds_j, in Table 3, along with other statistics. The means of the sideways deviations are equal to the cumulative sum of the forwards deviations, so start and end at 0. We show rows for the June values at the beginning and end of each year, for which the sideways deviations are necessarily zero.

3.16. We then show the standard deviations of the sideways deviations, and the T-ratios for the means, based on the standard errors of the means (not shown). These T-ratios are all quite reasonable, as compared with the forwards ones. The first sideways deviation is the same as the first forwards deviation, and the last pair are also the same, but with the sign changed. However, the standard errors are much bigger, and the large values for the monthly forwards deviations seem to balance out as the months go by. For the forwards deviations, four of the means exceed 0.2 (ignoring the sign); for the sideways ones only two do. The skewness and kurtosis of the sideways deviations (not shown) are also large, but no more so than these statistics for the forwards deviations.

3.17. We can divide each of the standard deviations by our estimate of σ_m of 0.006969. We show the result in the column headed "Multiple". We compare these with the values of rs_j for Brownian motion, shown in Table 2 of Part 3A, and repeated in Table 2 above. We then divide the multiple by rs_j to give an M-ratio. We expect these to be about around 1.0, but apart from the last two months, they are much bigger. One could, by simulation of true Brownian bridges, estimate the distribution of these M-ratios for Brownian bridges with 91 observations, but we have not done this, since it is clear that the data does not conform well to a theoretical Brownian bridge. But comparison of the values of the multiples in Table 3 with the values of rs_j for OU bridges in Table 1 of Part 3A shows that they are much bigger than for any OU model with a reasonable value of α_y.

3.18. When we look at the two periods separately, we find a big contrast. The first period shows M-ratios almost as big as the ones shown in Table 3 for the whole period. The second, on the other hand, shows M-ratios that are generally less than 1.0, only two being above 1.0, and most being between 0.8 and 0.9. It is possible that the lumpiness of the values in the early period allows them to bulge out further

https://doi.org/10.1017/S174849951600021X Published online by Cambridge University Press

than a true Brownian bridge would produce. For period (2B) the M-ratios are a little smaller than for (2A), so the excess M-ratios in period (1) are presumably not caused simply by the rounding.

3.19. A further calculation is for the within-year correlation coefficients between values of the forwards deviations, which we showed in section 2.8 of Part 3A for a Brownian bridge to be $-1/(n-1)$ or $1/11 = -0.0909$ in our case, the same for any pair of months. We can include only certain pairs for this calculation. For those one month apart, we have 11 pairs, but for those eleven months apart we have only one pair, so only 91 observations. Further, the large skewness and kurtosis would throw doubt on any tests of significance. The correlation coefficients are very unstable, ranging from +0.1259 for those 1 month apart to −0.3328 for those nine months apart.

3.20. When we look at the same within-year correlation coefficients for the two periods separately we see a similar effect, with very varied coefficients. They are slightly smaller for the second period, but the number of observations especially for those with large gaps, is getting rather small in both cases (only 33 at the minimum).

3.21. Another series of autocorrelation coefficients can be calculated, those for the whole series, treating it as continuous. These show high and very significant values, starting at +0.34 at lag 1, falling a little in the 1st year, but rising again to +0.50 at lag 12. The partial autocorrelation coefficients are also highly significant. This is indicative of strong connections between rates of inflation within years and in successive years, which we recognise in the autoregressive features of the annual model. These autocorrelation coefficients are rather smaller for the first period, but larger for the second, with +0.43 at lag 1, and +0.59 for lag 12. However, we return to this matter below.

3.22. A further feature can be observed. We have calculated the mean values and the standard deviations $\sigma_m(t)$ for each year (June to June). The different periods show different features. In the earlier years, the monthly mean (in effect one-twelfth of the rate of inflation in the year) is low, sometimes negative, and the standard deviations are high. In the 1970s and 1980s the rate of inflation is high, and the standard deviations are also high. In recent years, both have been low. We show this in Figure 12. We plot the monthly mean $I(t)/12$, for convenience of scaling.

3.23. This suggests that a function relating the monthly standard deviation to the annual rate of inflation could be introduced. This is far from uniform Brownian motion, but it does correspond to the concept of stochastic volatility and is quite like the ARCH (Autoregressive Conditional Heteroscedastic) model alternative for the Wilkie model. Rather than use the monthly mean we use the annual change in the year, which we denote $QLD(t) = QL(t+1) - QL(t)$. This equals $I(t+1)$. We use the value of $\sigma_m(t)$, calculated as described section 3.13, for months from year $t-1$ to year t.

3.24. We have calculated the linear regression of the estimated standard deviation, σ_m, and of the estimated annual variance, σ_m^2, against four functions of the value of $QLD(t)$, namely $\text{Abs}(QLD(t))$, $QLD(t)^2$, $\text{Abs}(QLD\ (t) - QSC)$ and $(QLD(t) - QSC)^2$, where QSC is the overall mean value of $QLD(t) = 0.041041$. This is all as described in section 8 of Part 3A. The correlation coefficients are shown in Table 4. We do the same calculations for the two sub-periods, and show the correlation coefficients for them too in Table 4. The values of QSC then differ; for 1923–1956 $QSC = 0.019169$, and for 1956–2014 $QSC = 0.053491$. In each case the regressions centred on zero are poorer, sometimes only by a little, than those centred on C. Sometimes the regressions for σ_m^2 are better than those for σ_m; and sometimes the reverse. Sometimes the regressions on the square are better than those on the absolute value; and sometimes the reverse.

https://doi.org/10.1017/S174849951600021X Published online by Cambridge University Press

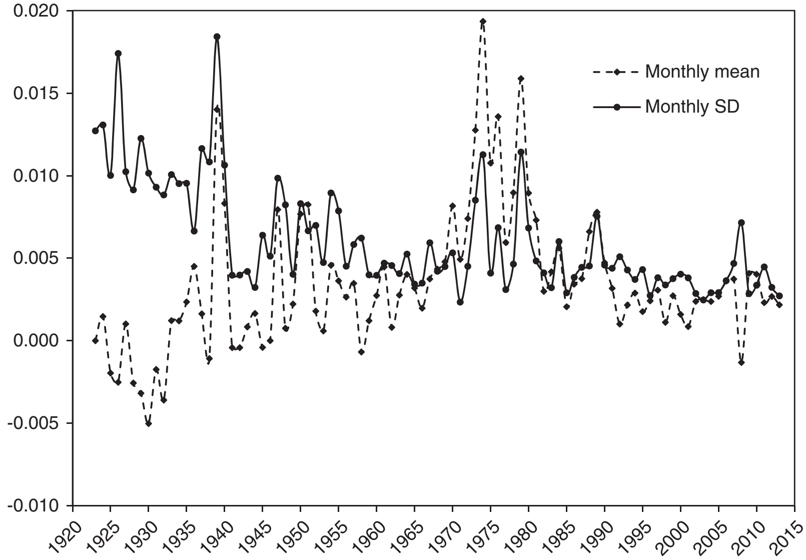

Figure 12. Means and standard deviations of Df_j in each year, from 1923 to 2014.

Table 4. Correlation coefficients from regressions, for different periods.

	1923–2014		1923–1956		1956–2014	
	σ_m	σ_m^2	σ_m	σ_m^2	σ_m	σ_m^2
Abs($QLD(t)$)	0.2713	0.2633	0.4051	0.4302	0.7282	0.7665
$QLD(t)^2$	0.3419	0.3345	0.4403	0.5133	0.7885	0.8553
Abs($QLD(t) - QSC$)	0.5865	0.5489	0.4650	0.4948	0.7733	0.8242
$(QLD(t) - QSC)^2$	0.4895	0.4715	0.5153	0.5897	0.7916	0.8741

3.25. We have now two plausible options. For the whole period the highest correlation coefficient is given by formula (5):

$$\sigma_m(t) = QSM(t) = QSA + QSB.\text{Abs}(QLD(t) - QSC)$$

$$\text{with } QSA = 0.004204,\ QSB = 0.055403 \text{ and } QSC = 0.041041$$

Our second choice for the whole period would be formula (6):

$$\sigma_m(t)^2 = QSM(t)^2 = QSA + QSB.\text{Abs}(QLD(t) - QSC)$$

$$\text{with } QSA = 0.00001689,\ QSB = 0.00091514 \text{ and } QSC \text{ again} = 0.041041$$

For the latter period alone, however, we could use formula (8):

$$\sigma_m(t)^2 = QSM(t)^2 = QSA + QSB.(QLD(t) - QSC)^2$$

with QSA = 0.00001570, QSB = 0.00410561 and QSC = 0.053491. In all cases the T-ratios of the parameters QSA and QSB are very large (between 6 and 14). We note that formula (6) gives the plot of $\sigma_m(t)$ against $QLD(t)$ as parts of two symmetrical parabolas, whereas formula (8) gives it as a

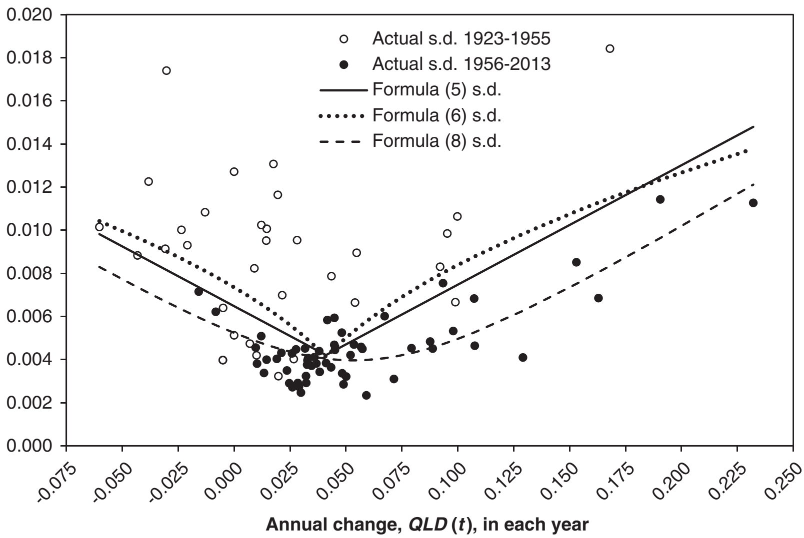

Figure 13. Standard deviations and three estimates of σ_m in each year, related to $QLD(t)$ in each year.

hyperbola, asymptotic to two straight lines as $QLD(t)$ goes to infinity in either direction. In this case formula (8) uses a different value of QSC so its minimum is offset.

3.26. In Figure 13, we plot actual and estimated values of σ_m against $I(t)$, using all the formulae (5), (6) and (8).

4. More Modelling of Retail Prices

4.1. We now have a plausible model for the standard deviations of the Brownian bridges we wish to apply to the values of $QL(t)$. We use Method 3 from section 2.10, but modified by the varying standard deviation. We can use either formula (5) or formula (8) from section 3.24 for this, and we call this Method 4. For simulations we use the initial conditions as at June 2014, and the actual values of $Q(t)$ for each month from June 2012 to June 2014. One simulation of values of $Q(t)$ for 50 years using formula (5) is shown in Figure 14 and corresponding values of $I(t)$ are shown in Figure 15. Visually, the results using formula (8) are very similar, but we find numerical differences, and we simulate for 250 years to get more reliable estimates. When we analyse the simulations in the same way as we have analysed the data, we find that the overall estimated value of σ_m for QL with formula (5) is 0.006423 and with formula (8) is 0.004984. In section 3.13 we observed that the estimated overall value of σ_m over the whole period was 0.006969, and over the period 1956–2014 was 0.004933, so the different results for the simulations are reasonably compatible with these.

4.2. We can compare the simulated values of $I(t)$ with the actual values from June 1964 to June 2014, also a period of 50 years, which are shown in Figure 16. We put Figures 15 and 16 on the same vertical scale to make comparison easier. While the overall patterns seem to us to be not unreasonable, the simulated values of $I(t)$ during each year seem to be much more jagged than the

https://doi.org/10.1017/S174849951600021X Published online by Cambridge University Press

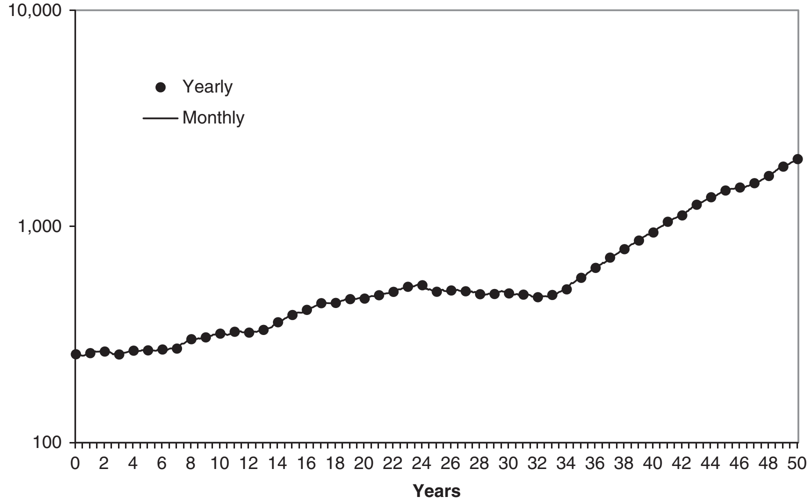

Figure 14. Stochastic yearly and monthly simulation for 50 years of $Q(t)$, Method 4, using formula (5), with initial conditions as at June 2014.

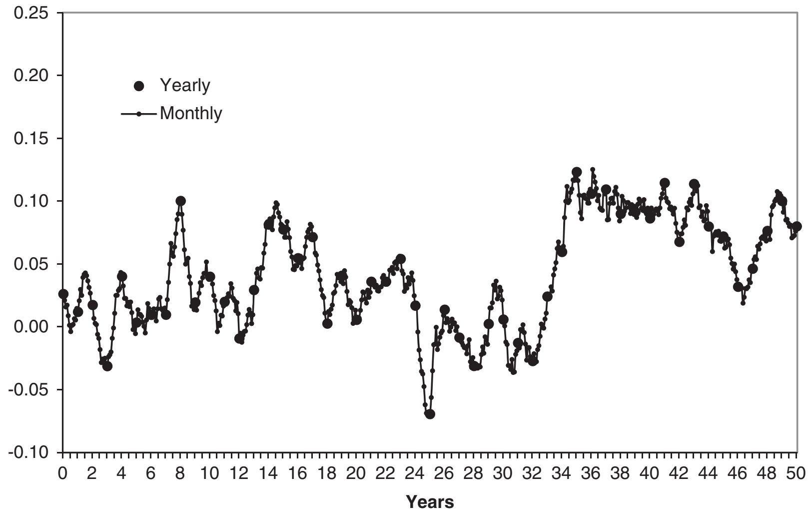

Figure 15. Stochastic yearly and monthly simulation for 50 years of $I(t)$, Method 4, using formula (5), with initial conditions as at June 2014.

actual values have been. We can confirm this numerically. We note that it is not the annual values that should be looked at, but the variation of monthly values within the year.

4.3. If we analyse the simulated values over 250 years of $I(t)$ in the same way as we have analysed $QL(t)$, we get a value for σ_m with formula (5) of 0.009065, which is 1.4113 time the value for $QL(t)$, equivalent to a ratio of the variances of 1.9917. When we use formula (8) we get a value for σ_m for $I(t)$ of 0.007041, corresponding with a ratio of the variances of 1.9958. We noted in section 2.10 that we would expect the values of $I(t)$ to have twice the variance of those of $QL(t)$, and the

https://doi.org/10.1017/S174849951600021X Published online by Cambridge University Press

simulated result is consistent with this. But for the actual data for 1923–2014 and for the two sub-periods we get the values shown in Table 5.

4.4. If we ignore the differences in the sizes of σ_m, we can see that the ratios of the values for $QL(t)$ and $I(t)$ are not replicated at all well in either simulation. The value of I in month j of year t is $I(t, j) = QL(t, j) - QL(t-1, j)$. The value one month later is $I(t, j+1) = QL(t, j+1) - QL(t-1, j+1)$. The change from $I(t, j)$ to $I(t, j+1)$ is equal to

$$\{QL(t, j+1) - QL(t, j)\} - \{QL(t-1, j+1) - QL(t-1, j)\}$$

The variances of these two terms are both σ_m^2 (though we note that they are in different years so they may be $\sigma_m(t)^2$ and $\sigma_m(t-1)^2$ if we use that model), so the variance of the change in $I(t)$ is $2\sigma_m^2$ unless the changes in the two years are correlated. However, if there is correlation of ρ between the changes (or between the Z innovations), then the variance of the change in $I(t)$ would be $2(1-\rho)\sigma_m^2$. To get the values down to those shown in the bottom line of Table 5, we need a value of ρ of about 0.45 for the whole.

4.5. We note that when we analyse the data we observe the equivalents of the Z^*s, but when simulating we need to use Zs. But provided the correlation coefficients for the calendar months have the same value, the Zs and the Z^*s have the same correlation coefficient.

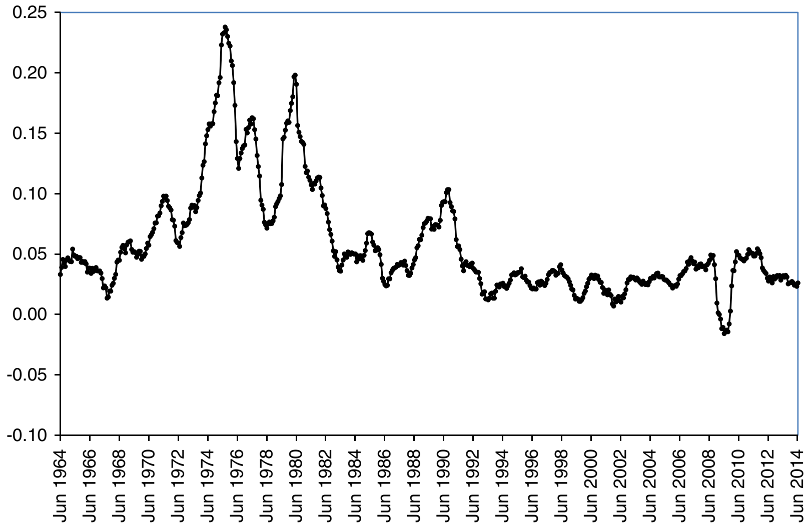

Figure 16. Values of $I(t)$ from June 1964 to June 2014.

Table 5. Values of σ_m for actual data for different periods, and for simulations for 250 years with different formulae.

	1923–2014	1923–1956	1956–2014	Simulation formula (5)	Simulation formula (8)
σ_m for $QL(t)$	0.006969	0.009548	0.004933	0.006423	0.004984
σ_m for $I(t)$	0.007338	0.010349	0.004851	0.009065	0.007041
Ratio	1.0529	1.0840	0.9835	1.4113	1.4127
Squared	1.1087	1.1750	0.9673	1.9917	1.9958

https://doi.org/10.1017/S174849951600021X Published online by Cambridge University Press

4.6. We observed in section 3.20 that if we treated the whole series of observed *QL*s as a uniform random walk, and calculated the autocorrelation coefficients, there was quite a large one of about 0.5 at lag 12. We now investigate this more closely. In each year t we take each of the standardised residuals, denoted $DfZ(t, j)$, as described in section 8 of Part 3A. We then, for each month j, calculate the correlation coefficient between the values of $DfZ(t, j)$ in successive years. Even though the overall correlation between years is taken care of in the annual model through the correlation between successive annual values of $I(t)$, there may still be correlation between the annual pattern of changes from month to month, as we had noticed with the large increases in the more recent period in April.

4.7. We show in Table 6 these 12-month correlation coefficients in each month for the whole period and each of the sub-periods. There is considerable variation from month to month, with some, such as indeed April, showing very high correlations, others rather low ones. The average over the whole period over all months is 0.5469, which is larger than we require to justify the values shown in Table 5.

4.8. We can now postulate Method 5, also based on Brownian bridges for *QL*, in which beside the standard deviations varying from year to year, we include also correlated innovations, *Z*s, so that each $Z(t, j)$ is correlated with the corresponding $Z(t-1, j)$ with correlation coefficient ρ, so that $Z(t, j) = \rho.\ Z(t-1, j) + \sqrt{(1-\rho^2)}.Z_2(t, j)$, where $Z_2(t, j)$ is another random unit innovation (normally distributed for the time being). We use the same value of ρ for each month, because we prefer not to have any part of the model specific to any particular calendar months. We believe that users would prefer this approach. We could also have introduced month-specific means for the innovations, but avoid this for the same reason. Further, the innovations, $Z(t, j)$ are not the same as the standardised forwards deviations, because of the deduction of $\sum_j Z(t, j)/n$. But if each monthly $Z(t, j)$ uses the same coefficient, the Zj^* terms and the forwards deviations are all connected with the same coefficient.

4.9. We now show in Figure 17 one simulation of values of $I(t)$ for 50 years, using Method 5 with formula (5) and with $\rho = 0.5469$. We compare this with Figure 16, considering not the annual pattern, but whether the monthly steps within each year seem to have similar properties, so that they might be

Table 6. Correlation coefficients between standardised forwards deviations in successive years for different periods.

Months	1923–2014	1923–1956	1956–2014
July	0.5746	0.2854	0.0750
August	0.1464	0.0710	0.1907
September	0.4643	0.3959	0.5337
October	0.2414	–0.0770	0.4488
November	0.4926	0.3502	0.3966
December	0.2298	–0.0179	0.3270
January	0.6102	0.2739	0.7051
February	0.4775	0.1724	0.4925
March	0.1325	–0.0555	0.2290
April	0.8076	0.6312	0.4457
May	0.1582	0.1126	0.2035
June	0.3702	0.3337	0.4461
Average	0.3921	0.2063	0.3745
Overall	0.5469	0.3886	0.6483

https://doi.org/10.1017/S174849951600021X Published online by Cambridge University Press

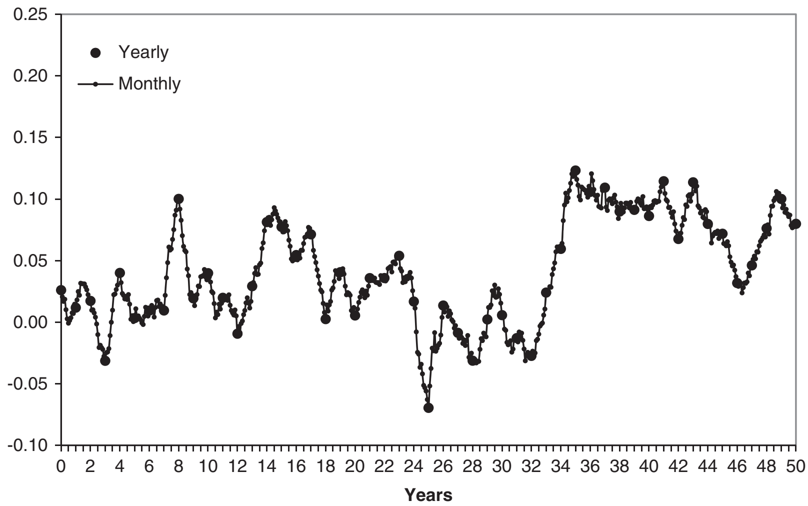

Figure 17. Stochastic yearly and monthly simulation for 50 years of $I(t)$, Method 5, formula (5) with $\rho = 0.5469$ and with initial conditions as at June 2014.

taken as two different sections of a much longer series. Whether Figure 20 on this criterion looks more like Figure 16 than Figure 15 does may be a matter of taste, but when we simulate over 250 years the values of σ_m are, for $QL(t)$ 0.006392, and for $I(t)$ 0.006298, giving a direct ratio of 0.9826 and a ratio of the variances of 0.9708. With a value of ρ of 0.5469 the theoretical ratio would have been 0.9062, so that the variance of changes in $I(t)$ would have been smaller than that for $QL(t)$ as was true for the period 1956–2014, but not for the whole period (see Table 5). Having nearly equal variances is not unreasonable, but it is only one simulation, albeit for 250 years.

4.10. We can summarise our potential models:

Model 1. OU bridge for $I(t)$; problem with scallops.

Model 2. Brownian bridge for $I(t)$; problem with increasing variance for QL.

Model 3. Brownian bridge for $QL(t)$: satisfactory shape, but more to come.

Model 4. As Model 3, with σ_m as a function of change in QL in the year.

Model 5. As Model 4, with correlation with previous year for innovations.

Model 5 seems altogether satisfactory.

4.11. We thus, after a significant amount of investigation, feel that we have a satisfactory way of stochastic interpolation for the inflation series. It is based on Brownian bridges for the $QL(t)$ series within each year, but with a standard deviation that depends on the rate of inflation in the year, and with innovations that have correlations across years for corresponding months. This result, along with those of all the other variables, is summarised in the Appendix of Part 3C, which forms a ·convenient reference. We now move on to the next series.

5. The Wages Model

5.1. The Wages model connects the Wages index, $W(t)$, its logarithm $WL(t)$ and the annual rate of wage increase, $J(t) = \ln W(t) - \ln W(t-1)$. $J(t)$ is dependent on the series for the annual rate of inflation, $I(t)$. The basic formulae for $J(t)$ and $W(t)$ are

$$WE(t) = WSD.WZ(t)$$

$$JN(t) = WA.JN(t-1) + WE(t)$$

$$J(t) = WW1.I(t) + WW2.I(t-1) + WMU + JN(t)$$

$$WL(t) = WL(t-1) + J(t)$$

$$W(t) = \exp(WL(t))$$

Two sets of parameters were suggested in Part 1 (Wilkie *et al.*, 2011), in both of which WA was set to 0, so although it might be an option, we shall ignore it in what follows. This, in effect puts $JN(t) = WE(t)$ either as

$$WW1 = 0.60,\ WW2 = 0.27,\ WMU = 0.020;\ WSD = 0.0219$$

in which $WW1 + WW2 \neq 1.0$ or else as

$$WW1 = 0.68,\ WW2 = 0.32,\ WMU = 0.015;\ WSD = 0.0228$$

in which $WW1 + WW2 = 1.0$.

5.2. The relationship between, $J(t)$, $WL(t)$ and $W(t)$ exactly parallels that between $I(t)$, $QL(t)$ and $Q(t)$, so we can immediately conclude that we should apply bridging to $WL(t)$ and not to $J(t)$ in the first place; instead the simulated monthly values of $J(t)$ will be derived from the simulated values of $WL(t)$. This follows at least Model 3 as used for $QL(t)$.

5.3. We can partition $J(t)$ into two parts, one depending on $I(t)$ and the other not, as

$$J1(t) = WW1.I(t) + WW2.I(t-1)$$

$$J2(t) = WMU + JN(t)$$

We can apply the partitioning to $WL(t)$ as

$$WL1(t) = WL1(t-1) + J1(t)$$

$$WL2(t) = WL2(t-1) + J2(t)$$

5.4. Since the values of $I(t)$ have, we can assume, already been simulated, we could then treat $J1(t)$ and $WL1(t)$ as already given and only simulate the remaining part, $WL2(t)$; or we could ignore the direct connection with price inflation, assuming that it is sufficiently taken into account in the (already simulated) annual model, and we could simulate monthly values of $WL(t)$ as a whole. We investigate both of these options in our analysis of the actual data.

6. The Wages Data Series

6.1. We have available what purport to be monthly values of the Wages index, in its many different forms, from January 1920. We show in Figure 18 values of the wages index, $W(t)$ (multiplied

https://doi.org/10.1017/S174849951600021X Published online by Cambridge University Press

by 2.5 to fit the scale better) from December 1922 to December 2014, which includes the period which we consider for analysis, and in Figure 19 values of the annual change in $W(t)$, $J(t)$, for the same period.

6.2. Until June 1934, we have available values only for June and December, and we have interpolated values for the intermediate months. While this leaves a series suitable for analysis at annual intervals, it is of no use for analysing the behaviour of a monthly series, so we omit it all, and start our effective data at June 1934. We then have a series with rather few significant figures, with a value of 34.5 initially, but rising to 167.2 by January 1968. During the earlier years the monthly differences lie on only a few possible lines, just as did the retail prices data.

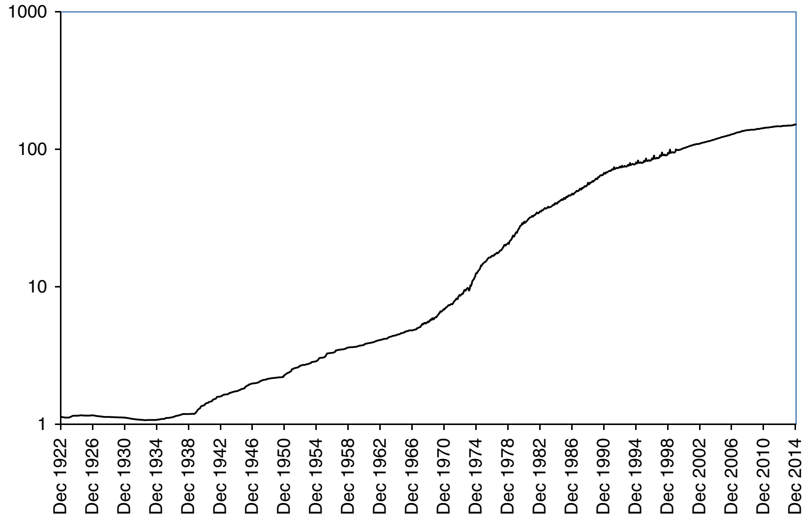

Figure 18. Values of the wages index, $W(t)$, monthly, from December 1922 to December 2014.

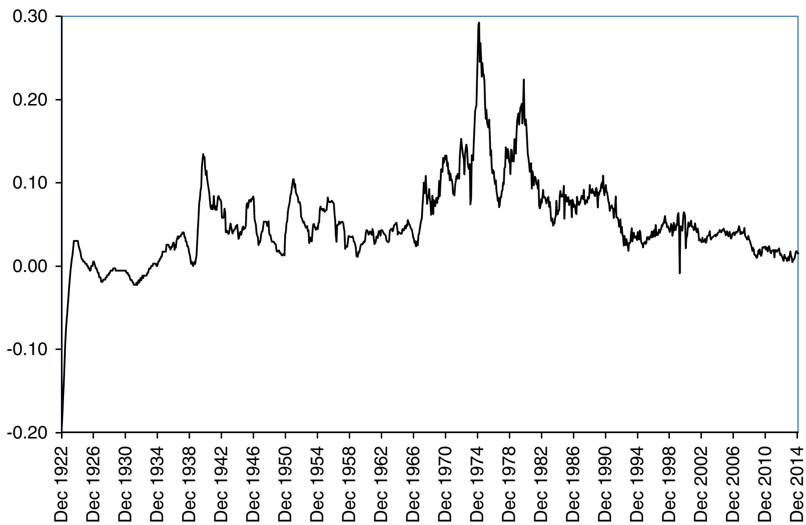

Figure 19. Values of $J(t)$, monthly, from December 1922 to December 2014.

https://doi.org/10.1017/S174849951600021X Published online by Cambridge University Press

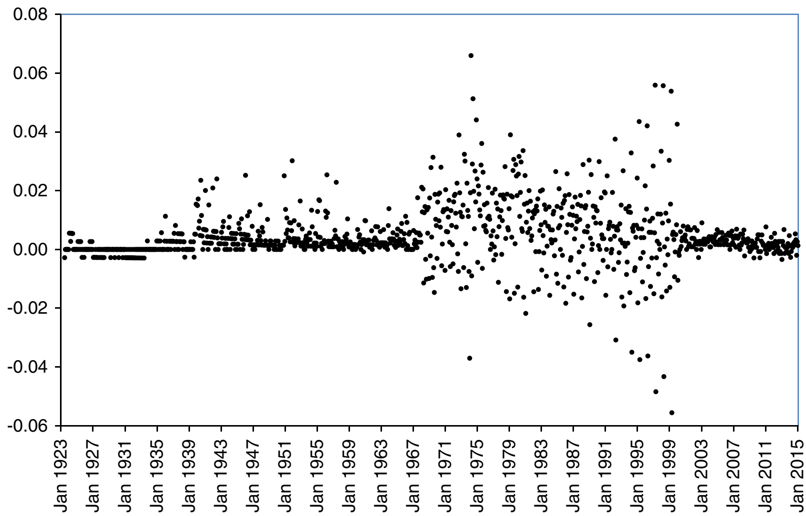

Figure 20. Monthly changes in $\ln W(t)$ from 1923 to 2014.

6.3. We show a chart of the monthly differences of the logarithms of $W(t)$ in Figure 20. This can be compared with the monthly differences of $\ln Q(t)$ shown in Figure 11. We see very small differences, often zero, up to 1934. Then we see changes lying on rather few curved lines, indicating changes of 0.1, 0.2,… in the basic series. In 1968, however, there is a big change, with the monthly differences fluctuating very much. We have had to use a series of different indices produced originally by the Department of Employment from January 1968 up to January 2000. From January 2000 we have used an index of "Average Weekly Earnings" (Series SA K54L), published by the Office of National Statistics, and we see a quite different pattern of much smaller monthly changes from 2000 onwards. We know that this last series is seasonally adjusted (SA). We know that the last few series we used prior to 2000 were not seasonally adjusted (NSA), and it is obvious that it is the difference between NSA and SA series that causes the change in pattern. The more scattered pattern is similar all the way from 1968, so presumably the earlier parts of this series were also NSA. We do not know whether the series from 1934 to 1968 was SA or not, but the pattern makes it look as if it had been.

6.4. We certainly need to analyse separately the four periods: (A) 1923–1934, (B) 1934–1968, (C) 1968–2000 and (D) 2000–2014. We also should analyse the whole period from 1934 to 2000 as well, omitting period (A). But we can also consider how to apply seasonal adjustment to our data.

6.5. The first, simple, comparison for the four periods is to look at the estimated values of σ_m, based on the forwards deviations. These are shown in Table 7. We denote the original, NSA, values for 1967–2000 as Period C1. Noting that the corresponding value for QL was 0.006969 we see that the value for C1 is much higher than this, and the values for the other three periods are much lower. It is worth considering how to seasonally adjust the 1967–2000 period.

6.6. For our seasonal adjustments we have first used a method based on our Brownian bridge analysis, but rather similar to certain traditional methods used by economists. Their method was to derive a "trend line" for the source data, possibly linear in $W(t)$, possibly exponential, i.e. linear in $WL(t)$, then derive average monthly differences from the trend line and then apply these average

https://doi.org/10.1017/S174849951600021X Published online by Cambridge University Press

Table 7. Estimated values of σ_m for WL for different periods.

Period	Time	Years	$\sigma_m(WL)$	$\sigma_m(QL)$
A	1923–1934	11	0.001505	0.011453
B	1934–1967	33	0.004426	0.007387
C1	1967–2000	33	0.015416	0.005422
D	2000–2014	14	0.002191	0.003832
All	1923–2014	91	0.009711	0.006969
BC1D	1934–2014	80	0.010342	0.006100

differences, either arithmetically or multiplicatively, to the source data to bring each month closer to the trend line.

6.7. We do something similar, but keep the June values of $WL(t)$ in each year unchanged, and fit a linear "trend line" to the annual data, which is simply the straight line of expected values for a Brownian bridge. We can then calculate the "sideways differences", analogous to sideways deviations, for each month, calculate the monthly average of these, and apply these averages to the observed values of $WL(t)$ to bring them (generally) closer to the trend line. This is equivalent to a multiplicative adjustment to each value of $W(t)$ (except for the June values).

6.8. In this first method we use the calculated sideways differences, but we observe for the Wages index as for the prices one that the monthly standard deviations of the forwards deviations seem to vary a lot from year to year depending on the mean change in that year. So a refinement is to divide each of the sideways differences by the estimated value of σ_m for the year, to give a scaled difference. We calculate the average of the scaled differences, then apply these, multiplied by the yearly estimate of σ_m to give what we hope will prove to be a reasonable SA value.

6.9. We do not show details, but we note that the increments in WL are relatively very large in March, June and November, and relatively very small in January, April and August. We would have expected January and April to be high, because these are the months following the end of the calendar year and the end of the tax year, when salary increases might be effective; but there may well be other quite different reasons, like the slacker period after Christmas, and the incidence of Easter, that cause what we observe.

6.10. We apply both these methods, and the resulting differences in $\ln W(t)$ in period C are somewhat improved, but still nothing like the SA values in period D. The value of σ_m is reduced, first to 0.011758 and then to 0.010715, both still relatively high. It is likely that those who construct proper SA series are able to take into account many other factors, such as the incidence of holidays (Easter may be in March or in April), the number of paydays in each month for weekly paid staff (each month has either four or five Fridays, but the pattern varies from year to year), and similar features. We have not gone so far as this, but we lay more stress on the SA series than on the NSA one.

6.11. We then compare the values of σ_m for WL with those for QL numerically. For Period B, the former is 0.60 times the latter; for Period D it is 0.57 times. To reduce the value for 1967–2000, even with the second adjustment we have used, to this sort of ratio would require reducing it by a factor of about 0.3. We try this, by taking the difference between the expected values of WL (from the linear

https://doi.org/10.1017/S174849951600021X Published online by Cambridge University Press

trend lines in each year) and the artificially "SA" values from our second method, and reducing that difference by a factor of 0.7, moving the value of WL towards the expected value. The resulting series give a value of σ_m of 0.003215, which is 0.6 times the corresponding value for QL. This is an ad hoc method with no theoretical justification, but it produces values of WL whose monthly differences, which are plotted in Figure 21, show reasonable similarity all the way along from 1934, except that they are rather larger in the 1970s and 1980s when annual wage increases, like annual price increases, were abnormally high. In Table 8, we show the resulting values of σ_m for Period C with different seasonal adjustments, which we denote C2, C3 and C4, and for the combined Periods B, C and D including different version for Period C. We continue with Period B, C4 and D.

6.12. At this stage, we can consider whether the splitting of WL into $WL1$ and $WL2$, as described in section 5.3, is worthwhile. The estimated value of σ_m for WL, for Period B, C4 and D (we use this period below without noting it again) is 0.003631. The corresponding value for $WL2$ is 0.021688, which is very much larger. Adjusting for the current rate of inflation as we go through the year does not seem to help at all. The calculation of $J1(t)$ depends on the annual rates of inflation $I(t)$ ending in month t and $I(t-12)$ ending in month $t-12$ and $WL1(t)$ is then an accumulation of values of $J1(t)$ at

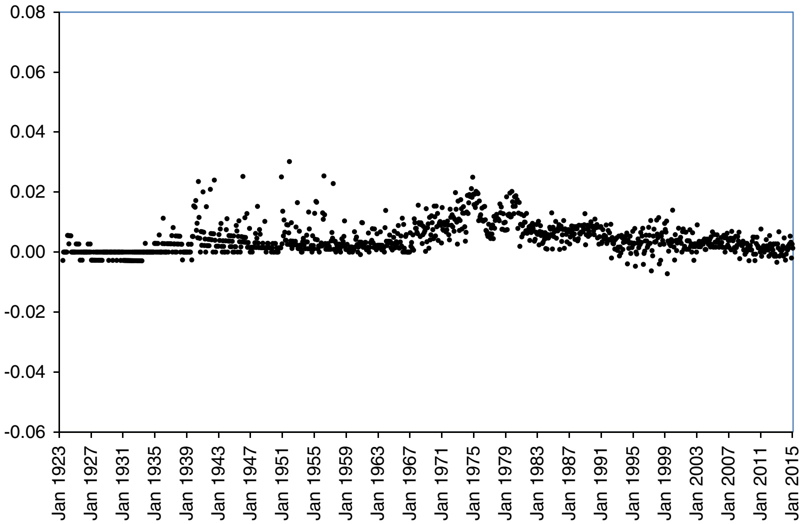

Figure 21. Monthly changes in $\ln W(t)$, adjusted with Series C4, from 1923 to 2014.

Table 8. Estimated values of σ_m for WL for different periods with seasonal adjustments.

Period	Time	Years	$\sigma_m(WL)$	$\sigma_m(QL)$
C1	1967–2000	33	0.015416	0.005422
C2	1967–2000	33	0.011758	0.005422
C3	1967–2000	33	0.010715	0.005422
C4	1967–2000	33	0.003215	0.005422
BC1D	1934–2014	80	0.010342	0.006100
BC2D	1934–2014	80	0.008121	0.006100
BC3D	1934–2014	80	0.007502	0.006100
BC4D	1934–2014	80	0.003631	0.006100

preceding yearly intervals. Any irregularity in the pattern of inflation, or in the way the series is started, is carried forward to all future years. Although it appears as if it should be a nice calculation, in practice it is difficult.

6.13. Nevertheless, the result is not surprising. Wage rates and salary rates within a firm have usually been adjusted once per year, even when inflation has been rather high. The negotiations or discussions leading up to a change may take into account the annual rate of price inflation over some recent period, but they cannot allow for the latest annual rate at the time the wage changes come into effect, because the relevant value of the price index would not be published till the month following, and probably the rate of wage changes would have been decided some time before that. The annual model for wage changes (the model for $J(t)$) takes into account proportions of the annual rates of price inflation for the current and the preceding year, but this represents a general association, rather than a precise month by month effect.

6.14. We therefore ignore $WL2$ and continue our investigations of WL. We note first that the overall value of σ_m is estimated at 0.003631, equivalent to an annual value, σ_y, of 0.0126. The corresponding variance from the annual model can be calculated as $(WW1.QSD)^2 + WSD^2$, which corresponds to a standard deviation of 0.03249, a very much bigger amount. This is a similar result to what we had for QL.

6.15. As for QL we can observe that when the annual change in WL is larger, there is some tendency for the standard deviation of the monthly increments to be also larger. We consider the same linear correlations as for QL, regressing either the equivalent annual estimates of the monthly standard deviation and the corresponding variance against various functions of the annual change in WL denoted $WLD(t) = J(t+1)$. Thus we regress σ_m^2 and σ_m against Abs($WLD(t)$), $WLD(t)^2$, Abs($WLD(t) - WSC$) and $(WLD(t) - WSC)^2$, where WSC is the overall mean value of $WLD(t) = 0.061688$. But we note that, since $WLD(t)$ is always positive, regressing against Abs($WLD(t)$) is the same as a linear regression against $WLD(t)$. The correlation coefficients are shown in Table 9.

6.16. These correlations are much weaker than those for price inflation, the highest correlation coefficient being 0.2826 for σ_m against $WLD(t)$; but we observe that when $WLD(t)$ was particularly high, as it was in the late 1970s and early 1980s, σ_m is about average, and the higher values of σ_m occur with values of $WLD(t)$ only a little above average. Experiments show, however, that if we use this regression, the values of $J(t)$ seem more plausible, as we shall show.

6.17. The best regression formula is a linear regression:

$$\sigma_m(t) = WSM(t) = WSA + WSB.WLD(t)$$

Table 9. Correlation coefficients from regressions.

	1934–2014	
	σ_m	σ_m^2
Abs($WLD(t)$)	0.2826	0.2451
$WLD(t)^2$	0.1475	0.1164
Abs($WLD(t) - WSC$)	−0.1358	−0.1410
$(WLD(t) - WSC)^2$	−0.0674	−0.0837

but we observe that there have been no negative values of $WLD(t)$ in our period, though there were large falls in wages in the early 1920s. If there were big falls again, we would expect a greater standard deviation within the year, so we modify the formula to be as formula (1):

$$\sigma_m(t) = WSM(t) = WSA + WSB.\text{Abs}(WLD(t))$$

with $WSA = 0.002586$ and $WSB = 0.010719$. The T-ratio of WSB is 2.6, so it is adequately significant.

6.18. The data points and the regression line are shown in Figure 22.

6.19. As we did with QL, we also look at the correlation between the standardised forwards deviations in the same months of successive years. For WL the correlations are weaker, with an overall correlation coefficient of 0.2047. This seems to be worth taking into account. We also observe that our estimated overall σ_m for $J(t)$ is 0.004310, which is quite a lot bigger than the overall σ_m for $WL(t)$ which we saw was 0.03631, so a correlation coefficient less than 0.5 is appropriate.

6.20. Finally, we can consider a new aspect, possible correlation between the forwards deviations of QL and WL in corresponding months, or months with small lags either way. We look at the forwards deviations as calculated, and at the standardised ones, each divided by the value of σ_m for the current year. The correlation coefficients, simultaneous and for small lags, are all small. As we have noted in section 6.12, the way wage increases are determined makes it difficult for the monthly increases in the price index to influence directly the corresponding monthly increases in the wage index, so such lack of direct correlation is not surprising.

7. Simulations for Wages

7.1. The analysis above gives us a model for simulating monthly values of WL which is on the same lines as that for QL, with a Brownian bridge in each year, but with the standard deviation varying

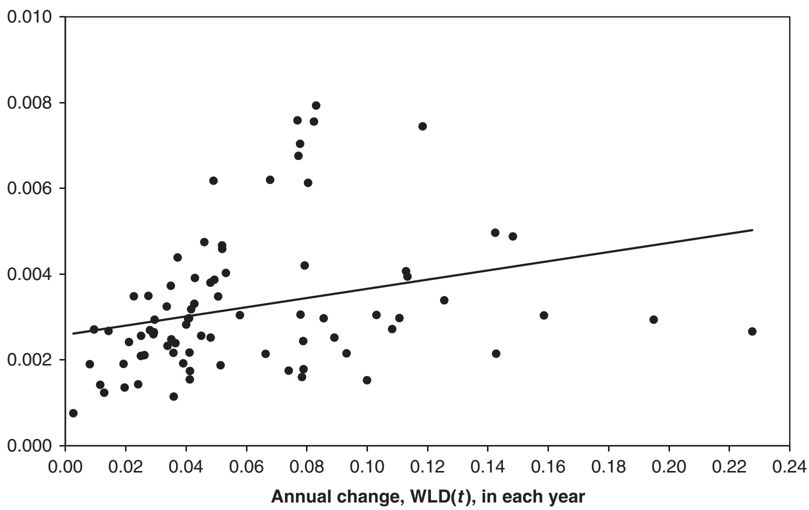

Figure 22. Standard deviations and estimates of σ_m in each year, related to $WLD(t)$ in each year, 1934–2014.

https://doi.org/10.1017/S174849951600021X Published online by Cambridge University Press

with the value of $J(t)$ in the year and with some correlation between the innovations for corresponding months in successive years.

7.2. In order to calculate simulated values also of $J(t)$, we need to supply values of WL for the year preceding our starting point, which we could do either by using the actual values, or by simulating them using a pure random walk.

7.3. In Figure 23 we show values of $J(t)$ for one simulation for 50 years, using the starting values as at June 2014. We believe that these look plausible in comparison with the actual series for $J(t)$ shown in Figure 24.

7.4. We do not feel wholly satisfied with our model for WL and J mainly because of the problems with the NSA data. However, if we consider the sort of situation in which actuaries might wish to simulate monthly for some contract, we would expect the monthly variation in the wages index to be unimportant. If one were simulating the results for a defined contribution pension plan for an individual, the contributions might vary with salary, but the salary would probably be assumed to be constant for a whole year. Even for a whole company, salaries would probably be assumed to change annually. The important part of monthly simulation is in the investment indices, which we discuss in Part 3C.

8. Conclusion

8.1. In this paper, we have investigated how it is suitable to apply stochastic bridging to the retail price series and the wages series of the Wilkie model, and have investigated the price and wages data over suitably long periods. We find complications in both the data series, but end up with what seems to be a satisfactory method of applying modified Brownian bridges to both series. In Part 3C we go

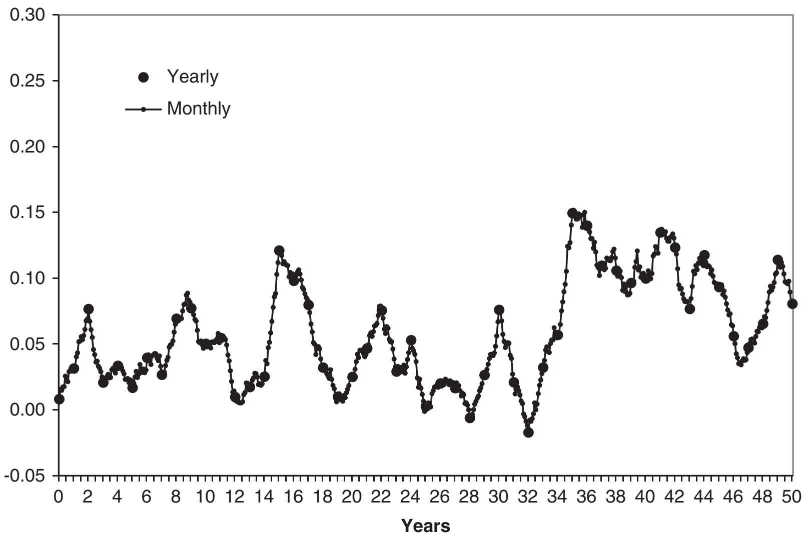

Figure 23. Stochastic yearly and monthly simulation for 50 years of $J(t)$, Method 5 of QL with $\rho = 0.2047$ and with initial conditions as at June 2014.

https://doi.org/10.1017/S174849951600021X Published online by Cambridge University Press

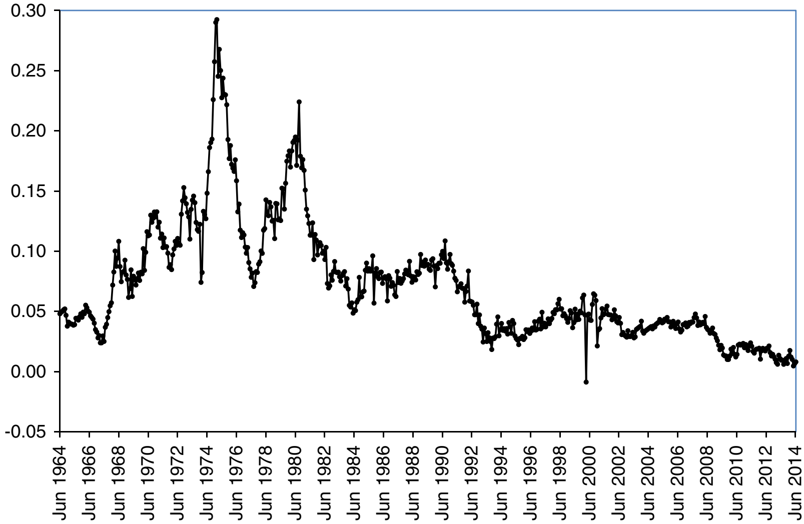

Figure 24. Values of $J(t)$, monthly from June 1964 to June 2014.

on investigating the other elements of the Wilkie model, share yields and dividends and interest rates, and draw some conclusions.

Acknowledgment

The authors are grateful to two anonymous referees for their comments.

References

Wilkie, A.D. (1986). A stochastic investment model for actuarial use. *Transactions of the Faculty of Actuaries*, **39**, 341–381.

Wilkie, A.D. (1995). More on a stochastic asset model for actuarial use. *British Actuarial Journal*, **1**, 777–964.

Wilkie, A.D. & Şahin, Ş. (2016). Yet more on a stochastic economic model: part 2: initial conditions, select periods and neutralising parameters. *Annals of Actuarial Science*, **10**, 1–51.

Wilkie, A.D., Şahin, Ş., Cairns, A.J.G. & Kleinow, T. (2011). Yet more on a stochastic economic model: part 1: updating and refitting, 1995 to 2009. *Annals of Actuarial Science*, **5**, 53–99.